

Effets des variations démographiques sur la structure génétique de populations, dans le cadre d'une maladie émergente

Méline Saubin

▶ **To cite this version:**

Méline Saubin. Effets des variations démographiques sur la structure génétique de populations, dans le cadre d'une maladie émergente. Génétique des populations [q-bio.PE]. AgroParisTech, 2022. Français. ⟨NNT : 2022AGPT0013⟩. ⟨tel-03954624⟩

HAL Id: tel-03954624

https://pastel.hal.science/tel-03954624v1

Submitted on 24 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

NNT: 2022AGPT0013

THESE DE DOCTORAT

pour obtenir le grade de

Docteur d'AgroParisTech

Spécialité : Biologie et écologie des forêts et des agrosystèmes

École doctorale n°607
SCIENCE ET INGENIERIE DES RESSOURCES NATURELLES (SIReNa)

par

Méline SAUBIN

Effets des variations démographiques sur la structure génétique de populations, dans le cadre d'une maladie émergente

Directeur de thèse : **Pascal FREY**
Co-encadrant de thèse : **Fabien HALKETT**

Thèse présentée et soutenue à Champenoux, le 8 décembre 2022

Composition du jury :

Virginie RAVIGNÉ, Chercheuse (HDR), CIRAD (Montpellier)	Présidente & Rapporteur
Maria ORIVE, Professeure, University of Kansas (Etats-Unis)	Rapporteur & Examinatrice
Myriam VALÉRO, Directrice de recherche, CNRS (Roscoff)	Rapporteur & Examinatrice
Thomas GUILLEMAUD, Directeur de recherche, INRAE (centre PACA)	Examinateur
Julien PAPAÏX, Chargé de recherche, INRAE (centre PACA)	Examinateur
Pascal FREY, Directeur de recherche, INRAE (centre Grand-Est-Nancy)	Directeur de thèse
Fabien HALKETT, Chargé de recherche, INRAE (centre Grand-Est-Nancy)	Co-encadrant de thèse

UMR 1136 INRAE / Université de Lorraine, Interactions Arbres - Microorganismes
INRAE Centre Grand Est - Nancy
54280 Champenoux

Remerciements

Je remercie tout d'abord les membres de mon jury de thèse **Maria Orive**, **Virginie Ravigné**, **Myriam Valéro**, **Julien Papaïx** et **Thomas Guillemaud** pour avoir accepté d'évaluer mon travail de thèse.

Je souhaite ensuite exprimer ma plus profonde reconnaissance à mes encadrants de thèse **Fabien Halkett** et **Pascal Frey**, qui m'ont accompagnée pendant trois ans et m'ont permis de réaliser ce travail jusqu'au bout dans le meilleur des cadres possibles. Fabien, je te remercie particulièrement pour ta présence et ton implication depuis le premier jour de mon arrivée en stage de M2 puis tout au long de ma thèse. Je te remercie pour tes conseils, ta patience lors de nos séances de co-rédaction d'articles et enfin pour le ravitaillement en thé et en raisins secs durant les différentes étapes de la rédaction. Tu m'as donné l'envie de me poser tout un tas de questions et surtout de tout faire pour y répondre. Pascal, je te remercie beaucoup pour ton aide durant cette thèse, pour tes réponses tellement riches à toutes mes questions sur Mlp, les agents pathogènes, leurs cycles de vie et la phytopathologie en général. Je te remercie enfin pour ta bonne humeur à toute épreuve et ta volonté de transmettre. Un grand merci à vous deux pour tout ce que vous avez fait pour moi durant ces trois ans.

Je remercie les membres de mon comité de thèse qui m'ont suivie et ont tenté de me rendre plus réaliste sur le calendrier : **Benoît Marçais**, **Frédéric Fabre**, **Frédéric Grognard**, **Lydia Bousset** et **Solenn Stoeckel**.

Je remercie mes collaborateurs et collaboratrices aux quatre coins de la France : **Suzanne Touzeau**, **Clotilde Djuikem** et **Frédéric Grognard** à Antibes, **Jérôme Coville** à Avignon, **Frédéric Fabre** à Bordeaux et enfin **Ronan Becheler** et **Solenn Stoeckel** à Rennes. Un grand merci à Frédéric F. et Solenn, vous m'avez tous les deux accompagnée, chacun dans une partie distincte de ma thèse, et m'avez permis d'acquérir une formidable ouverture d'esprit. Merci pour votre patience, votre bienveillance, nos discussions en visio et par téléphone, merci de m'avoir aidée comme vous l'avez fait malgré la distance.

Je remercie mes collaborateurs au-delà des frontières françaises : **Bruno Sudret** et **Xujia Zhu** à Zurich et bien sûr **Aurélien Tellier** à Freising. Aurélien, je te remercie particulièrement de m'avoir accueillie un mois dans ton équipe et de m'avoir si bien accompa-

gnée pendant une partie de ma thèse.

Je remercie l'ensemble de l'équipe pathologie, toutes ces personnes formidables avec qui j'ai passé ces trois belles années. **Jérémy**, **Bénédicte**, **Axelle**, **Anaïs**, merci pour vos discussions, vos revendications, votre humour et votre dynamisme. Merci bien sûr pour toutes ces pauses-café qui m'ont tellement souvent permis de prendre du recul. Merci à **Clémentine** et **Simon** avec qui j'ai partagé mon bureau pendant une grande partie de ma thèse. Merci pour votre bonne humeur, vos questions et vos réponses, nos conversations et les petites pauses bienvenues. Mille mercis en particulier à Clémentine qui a été particulièrement présente pendant ma thèse et m'a permis de me rendre compte un an à l'avance de ce qui m'attendait à chaque étape. Merci pour ta bonne humeur, ton enthousiasme, ton aide et nos soirées au bar. Je remercie également **Benoît**, **Christophe**, **Claude**, **Emma**, **Léane**, **Mireia** et **Olivier**.

Merci à **Fadwa** et **Gaurav** pour les excellents moments que l'on a passés ensemble en attendant le bus.

Je remercie mes amis sur Nancy : **Claire**, **Emmanuel**, et **Philippe**. Merci pour toutes ces soirées passées ensemble, merci de m'avoir écoutée râler et m'avoir souvent changé les idées. Je remercie mes amis hors de Nancy, sur Discord, en Ardèche, à Valence, Grenoble, Paris et ailleurs, qui ont chacun participé à leur manière à rendre ces dernières années exceptionnelles. Une mention spéciale pour **Nico** et **JP**, pour toutes ces soirées passées ensemble malgré la distance et malgré le fait que nous ne vivions pas sur le même fuseau horaire du sommeil.

Un remerciement particulier à ma famille, qui a depuis toujours le privilège d'être sollicitée dans les moments les plus difficiles. **Papa**, **Maman**, **Élodie** et **Johan**, merci à tous les quatre pour votre soutien sans faille et pour ces confinements inoubliables. Élodie et Maman, un grand merci pour vos relectures de ma thèse et pour m'avoir fait réaliser, mieux vaut tard que jamais, qu'il existe sans doute des règles grammaticales précises pour le placement des virgules dans une phrase.

Et pour finir, **Antoine**. Merci pour ta bonne humeur et pour ton soutien moral permanent pendant ces trois ans. Merci pour absolument tout ce que tu es et pour tout ce que tu représentes. Merci à toute ta famille, **Denis**, **Nathalie**, **Susie** et **Manon** pour votre soutien et les superbes week-ends passés en votre compagnie.

Un grand merci à tous et surtout, si cette thèse était à refaire dans ces conditions, je la referais sans hésiter.

Table des matières

Remerciements iii

Table des matières v

Liste des figures vii

Liste des tableaux viii

Acronymes et anglicismes ix

Introduction générale 1
Les maladies émergentes . . . 3
Évolution et génétique des populations . . . 4
Apports de la démogénétique . . . 7
De la modélisation à l'inférence . . . 9
Modèle d'étude : *Melampsora larici-populina*, agent de la rouille du peuplier . . . 11
Problématiques et objectifs . . . 14

I Suivi temporel d'une maladie ré-émergente : le cas d'un contournement de résistance 17

Chapitre 1 Introduction 19
1.1 Les contournements de résistances . . . 21
1.2 Article 1. Improving sustainable crop protection using population genetics concepts . . . 26
1.3 Cas d'étude : le contournement de la résistance RMlp7 par Mlp . . . 61
1.4 Objectifs . . . 64

Chapitre 2 Étude épidemio-évolutive d'un contournement de résistance 67
2.1 Article 2. Impact of ploidy and pathogen life cycle on resistance durability . . . 69
2.2 Analyse déterministe par résolution analytique . . . 101
2.2.1 Enjeux de la résolution déterministe . . . 101
2.2.2 Description du modèle . . . 101
2.2.3 Les différents équilibres et leur stabilité . . . 104
2.2.3.1 Extinction de la population . . . 105
2.2.3.2 Disparition des individus hétérozygotes Aa . . . 107

2.2.3.3 Disparition des individus homozygotes AA . . . 109
2.2.3.4 Fixation de l'allèle A . . . 111
2.2.3.5 Fixation de l'allèle a . . . 114
2.2.3.6 Survie des individus hétérozygotes Aa uniquement . . . 115
2.2.3.7 Conditions restantes . . . 116
2.2.4 Conclusion . . . 119
2.3 Conclusion en regard du contournement de la résistance RMlp7 par Mlp . . . 119

Chapitre 3 Etude démogénétique d'un contournement de résistance 123
3.1 Article 3. Interplay between demography and selection: a forward model illuminates the temporal genetic signatures of rapid adaptation in plant pathogens 125
3.2 Conclusion en regard du contournement de la résistance RMlp7 par Mlp . . . 157

Chapitre 4 Inférence d'un scénario de contournement 161
4.1 Article 4. Approximate Bayesian Computation applied to time series of population genetic data disentangles rapid genetic changes and demographic variations . . . 163
4.2 Inférence du scénario de contournement de la résistance Rmlp7 . . . 193

Conclusion de la partie I 195

II Étude d'une maladie ré-émergente en système spatialisé : le cas d'une colonisation clonale 197

Chapitre 5 Introduction 199
5.1 Dispersion et démographie . . . 201
5.2 Structures génétiques sur un front de colonisation . . . 203
5.3 Cas d'étude : les invasions récurrentes de Mlp le long de la vallée de la Durance 206
5.4 Objectifs . . . 207

Chapitre 6 Modèle mécanico-statistique et inférence de la dynamique d'expansion de Mlp dans la vallée de la Durance 209
6.1 Article 5. Inferring invasion determinants with mechanistic models and multitype samples . . . 211
6.2 Dynamique d'expansion de Mlp dans la vallée de la Durance . . . 246

Chapitre 7 Variabilité inter-annuelle de la structure génétique de Mlp dans la vallée de la Durance 249
7.1 Article 6. Genetic signatures of a range expansion *in natura*: variation across years . . . 251
7.2 Dynamique d'expansion et structures génétiques . . . 272

Conclusion de la partie II 273

Conclusion générale et perspectives 275

Bibliographie 281

Annexes 289

Liste des figures

1 Schéma des principales forces évolutives 4
2 Schéma du lien entre structure génétique et démographie 8
3 Schéma de l'estimation d'un paramètre θ par ABC 11
4 Cycle de vie de *Melampsora larici-populina* 13
5 Photographies des écidies et des urédies de Mlp sur mélèze et peuplier 14
6 Photographies des différents impacts de Mlp sur les peupleraies cultivées françaises 14

1.1 Schéma d'un contournement de résistance 22
1.2 Schéma du modèle gène-pour-gène selon la ploïdie 23
1.3 Les cycles boom-and-bust 25
1.4 Contournement de la résistance RMlp7 par Mlp 62
1.5 Evolution de la structure génétique de Mlp 63
1.6 Schéma du modèle gène-pour-gène appliqué à Mlp 64

3.1 Évolution empirique de quatre indices de génétique des populations chez Mlp 159

5.1 Schéma des quatre principaux mécanismes écologiques impactant la dynamique d'une population sur un front de colonisation 201
5.2 Fonctions de densité de trois noyaux de dispersion 202
5.3 Schéma des quatre principaux mécanismes évolutifs impactant la structure génétique neutre d'une population sur un front de colonisation 203
5.4 Évolution de la diversité génétique le long d'un front de colonisation 205
5.5 Évolution de la diversité génétique le long d'un front de colonisation selon la fréquence de LDD 206

7.1 Schéma du lien entre génétique des populations et épidémiologie 278

Liste des tableaux

2.1 Description des paramètres du modèle 102
2.2 Tableau bilan de l'ensemble des équilibres étudiés 118

Acronymes et anglicismes

ABC Approximate Bayesian Computation (Calcul Bayésien Approché)

LDD Long-Distance Dispersal (Dispersion à Longue Distance)

Mlp *Melampsora larici-populina* (Agent pathogène responsable de la maladie de la rouille du peuplier)

SNP Single Nucleotide Polymorphism (Polymorphisme Nucléotidique Ponctuel)

INTRODUCTION GÉNÉRALE

Les maladies émergentes

Une maladie émergente peut-être définie comme une « maladie dont l'incidence réelle augmente de manière significative dans une population donnée, d'une région donnée et pendant une période donnée, par rapport à la situation épidémiologique habituelle de cette maladie » (TOMA ET THIRY, 2003). Les maladies émergentes peuvent alors être causées par l'évolution d'agents pathogènes déjà existants (on parle alors d'émergence réelle), ou dont l'aire géographique augmente rapidement sur un territoire précédemment non affecté (on parle alors d'émergence géographique) (MOUTOU ET PASTORET, 2015). Ces maladies peuvent avoir des conséquences importantes sur les écosystèmes et peuvent profondément modifier la diversité génétique de certaines espèces. La compréhension des maladies émergentes est primordiale en écologie et fait l'objet de nombreuses études, visant notamment à prédire l'évolution spatio-temporelle des populations d'agents pathogènes (KOT *et al.*, 1996).

Dans les agrosystèmes en particulier, l'évolution particulièrement rapide des agents pathogènes (PERKINS *et al.*, 2013) et la perte de diversité génétique des hôtes induite par l'agriculture peuvent favoriser et accélérer l'expansion de telles épidémies (MUNDT, 2002; GILLIGAN ET VAN DEN BOSCH, 2008; SAPOUKHINA *et al.*, 2009; BURDON *et al.*, 2014; DJIAN-CAPORALINO *et al.*, 2014; FABRE *et al.*, 2015; BOUSSET *et al.*, 2018; RIMBAUD *et al.*, 2018). Le développement de plants génétiquement résistants aux pathogènes est une stratégie largement utilisée dans le monde agricole. Le déploiement d'une résistance dans le paysage est souvent suivi, plus ou moins rapidement, du contournement de cette résistance lors de l'adaptation de l'agent pathogène (JOHNSON, 1984; PINK ET PUDDEPHAT, 1999; BROWN ET TELLIER, 2011; BURDON *et al.*, 2016). Ces contournements de résistance successifs résultent de l'évolution d'agents pathogènes déjà présents dans l'écosystème. Ce sont des exemples classiques de maladies émergentes réelles.

D'un autre côté, pour mieux comprendre l'expansion des agents pathogènes et prédire la colonisation de nouveaux environnements il faut étudier finement l'ensemble des processus de dispersion. La mondialisation et l'intensification des pratiques agricoles participent à leur introduction sur de nouvelles aires géographiques, ce qui accentue le nombre et la rapidité des émergences (HALKETT *et al.*, 2021; ANDERSON *et al.*, 2004). Ces maladies émergentes peuvent mener à des épidémies majeures, comme l'illustrent deux exemples tristement célèbres pour leurs conséquences épidémiques : la maladie de la pyriculariose du blé qui a émergé en 2016 et a dévasté 16% de la surface cultivée dans les champs de blé du Bangladesh (ISLAM *et al.*, 2016) ; et l'émergence du champignon ascomycète *Hymenoscyphus fraxineus* infectant les frênes sur l'ensemble du continent européen (MARÇAIS ET HUSSON, 2017). Il s'agit dans ce cas d'émergences géographiques.

Évolution et génétique des populations

La notion de population définit un ensemble d'individus appartenant à la même espèce, partageant une même localisation, et capables de se reproduire entre eux. Si cette définition est communément admise dans les études d'écologie et d'évolution, la taille d'une population et sa délimitation dans l'espace peuvent cependant varier considérablement d'une étude à l'autre, selon l'échelle considérée.

Différentes forces évolutives s'appliquent aux populations et impactent leur évolution génétique (FIGURE 1). Ces forces évolutives peuvent s'appliquer à différentes échelles : une séquence nucléotidique, un allèle, un ensemble de gènes, un individu, une population, ou un ensemble de populations.

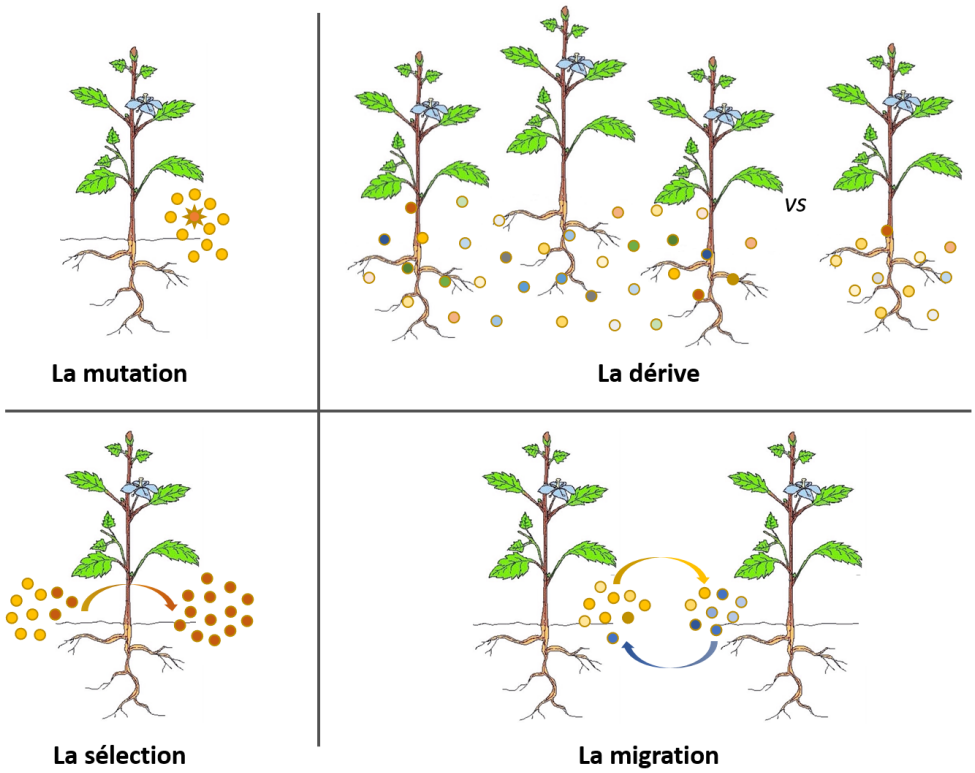


FIGURE 1 – Schéma des principales forces évolutives impactant l'évolution d'une population (d'après GÓMEZ *et al.*, 2009 et HALKETT *et al.*, 2021). Les ronds représentent les agents pathogènes, les couleurs représentent leurs fonds génétiques.

- **La mutation** est un changement aléatoire de la séquence nucléotidique localisée dans le génome d'un individu. La mutation est donc une source de variabilité génétique au sein des individus, qui se traduit par une variabilité génétique au sein des populations via l'apparition aléatoire de variants génétiques. Si la mutation est ponctuelle, elle crée un nouvel allèle. Une mutation dans une région codante du génome peut alors affecter la fonction d'un gène. Ce mécanisme est le principal déterminant de l'adaptation d'un

agent pathogène qui conduit notamment aux contournements des résistances des plantes (HALKETT *et al.*, 2021).

- **La recombinaison** est un mécanisme lié à la reproduction sexuée, qui conduit chez un descendant à une association allélique différente de celles observées chez les parents. Elle génère une diversité génotypique.
- **La dérive génétique** est la variation aléatoire des fréquences alléliques au cours du temps dans une population de taille finie, de par la transmission au hasard des allèles au cours des générations. L'effet de la dérive est d'autant plus important que la population est petite. Elle peut mener par hasard à la fixation ou à la disparition d'un allèle, ce quelle que soit la valeur sélective de cet allèle. La dérive entraîne donc une diminution de la diversité génétique chez les populations de petite taille ou subissant une réduction de taille durant leur cycle de vie (par exemple lors de la survie hivernale).
- **La sélection** est la variation des fréquences alléliques dans une population due aux différences de survie et de reproduction entre individus. Une sélection positive ou unidirectionelle entraîne l'augmentation en fréquence des allèles conférant un avantage sélectif. Les allèles concernés sont ceux qui, dans un environnement donné, entraînent une augmentation du nombre moyen de descendants chez les individus qui les portent. Il s'agit du processus menant à l'adaptation des populations aux conditions de leur milieu de vie.
- **La migration** représente les déplacements d'individus entre populations. Ces échanges d'individus (et donc des allèles qu'ils portent) se traduisent par des variations génétiques, ou flux de gènes, dans les populations concernées. La migration peut entraîner un gain ou une perte de diversité génétique selon le sens du flux migratoire. Elle impacte également le nombre d'individus d'une population et donc la susceptibilité de cette population à la dérive génétique. Entre deux populations établies échangeant des migrants, la dérive est connue pour homogénéiser les diversités génétiques. A l'inverse, lors de la colonisation d'un nouveau milieu, la migration crée une différentiation génétique entre la population source (dont sont issus les individus migrants) et la population nouvellement fondée par un faible nombre d'individus. Cette différenciation génétique liée à la migration est appelée l'effet de fondation.

L'étude de l'ensemble de ces variations de fréquences alléliques au sein et entre populations est au coeur de la génétique des populations (Encadré 1). Cette discipline étudie la variabilité génétique et son partitionnement intra et inter populations. Elle permet notamment de définir des populations *a posteriori*, d'après le partitionnement de leur variabilité génétique. L'étude de ces structures génétiques permet de comprendre les effets relatifs de chacune des forces évolutives précédemment décrites et de retracer l'histoire évolutive d'une population.

Encadré 1 - Les indices de génétique des populations

De nombreux indices de génétique des populations ont été décrits et peuvent être regroupés en plusieurs catégories, selon le niveau d'organisation auxquels ils s'intéressent :

1. Les indices intra-populationnels

Ils sont utilisés pour décrire la variabilité génétique au sein d'une population. Parmi les plus utilisés, nous pouvons distinguer dans un premier temps les indices qui évaluent la variation à un locus :

- **L'hétérozygotie observée** H_O représente la proportion réelle d'individus hétérozygotes aux loci étudiés.
- **L'hétérozygotie théorique** H_E représente la proportion d'individus hétérozygotes aux loci étudiés sous l'hypothèse d'une population à l'équilibre de Hardy-Weinberg, *i.e.* une population panmictique, pangamique, de grande taille, isolée et sans mutation ni sélection. Elle correspond à la probabilité de tirer deux allèles différents au hasard, et correspond donc à une mesure de diversité génétique. Il s'agit d'un indice robuste et non biaisé par la taille de l'échantillonnage (NEI, 1978).
- **L'indice de fixation** F_{IS} aussi appelé coefficient de consanguinité (WRIGHT, 1931, 1949).
  $F_{IS} = 1 - \frac{H_O}{H_E}$ et mesure le partitionnement entre la variabilité génétique intra-individus et celle inter-individus au sein d'une population. Cela revient à mesurer l'écart à la panmixie (excès d'homozygotes) au sein de la population. Ainsi, $F_{IS} = 0$ lorsque la population est à l'équilibre de Hardy-Weinberg, $0 < F_{IS} < 1$ lorsque la population présente un excès d'homozygotes, et $F_{IS} = 1$ en cas de fixation complète de l'homozygotie.

Parmi les indice intra-populationnels, nous pouvons distinguer dans un second temps les indices qui évaluent la variation à plusieurs loci :

- **Le déséquilibre de liaison** LD mesure la différence d'associations entre les allèles de deux loci dans la population par rapport à l'association obtenue si les allèles de ces loci étaient tirés au hasard. Ce déséquilibre nous renseigne sur l'indépendance statistique ($LD = 0$) ou non ($LD > 0$) entre les génotypes de deux loci. Pour évaluer le déséquilibre de liaison, il existe un indice combiné multilocus, $\bar{r}_D$, qui est débiaisé par rapport au nombre de loci. Cependant, cet indice reste très sensible à la taille de l'échantillonnage et doit souvent être interprété avec précaution (DE MEEÛS ET BALLOUX, 2004).
- **La richesse génotypique** $R = \frac{G-1}{N-1}$ représente le ratio entre le nombre de génotypes distincts G dans la population et la taille de la population N (ARNAUD-HAOND *et al.*, 2007).

Encadré 1 - Les indices de génétique des populations

2. Les indices inter-populationnels

Ils sont utilisés pour décrire la variabilité génétique entre populations. Le plus utilisé est l'indice de fixation F_{ST} qui mesure le degré de différenciation entre populations (WRIGHT, 1949; WEIR ET COCKERHAM, 1984). Il est défini comme la proportion de la variation génétique totale expliquée par la variation génétique entre les sous-populations, et traduit donc la diminution de l'hétérozygotie liée à la divergence entre une sous-population et la population totale. Il s'exprime en fonction du F_{IS} (différentiation des individus au sein des populations) et du F_{IT} (différenciation des individus par rapport au total) :

$$F_{ST} = 1 - \frac{1 - F_{IS}}{1 - F_{IT}}$$

D'après WRIGHT (1978) :
$0 < F_{ST} < 0.05$ révèle une faible différenciation,
$0.05 < F_{ST} < 0.15$ une différenciation modérée,
$0.15 < F_{ST} < 0.25$ une différenciation importante, et
$0.25 < F_{ST}$ une différenciation très importante.

Apports de la démogénétique

Les variations de taille de populations et les structures génétiques sont étroitement liées, *via* trois forces évolutives précédemment décrites : la dérive génétique, la sélection et la migration. Deux méthodes distinctes sont principalement utilisées pour inférer les paramètres démographiques (la taille de population, la dispersion, le mode de reproduction, etc.) d'une population d'agents pathogènes : 1) la méthode directe avec un suivi de la dynamique des populations par des données de comptage ; et 2) la méthode dite indirecte qui consiste à utiliser les outils de génétique des populations, *via* des formules liant les indices de diversité génétique aux paramètres démographiques à inférer (FIGURE 2). Ces formules permettent par exemple de lier directement la différenciation génétique d'une population à sa taille et à son taux de migration ($F_{ST} \approx \frac{1}{4Nm}$ avec N la taille de population, et m le taux de migration, WRIGHT, 1931). L'utilisation de ces formules pour relier génétique et démographie est cependant conditionnée par une forte hypothèse qui est de considérer que l'équilibre démographique est atteint dans la population, avec des fréquences alléliques et génotypiques stables d'une génération à l'autre. Or, dans la plupart des situations, cette hypothèse est loin d'être vérifiée et ces formules ne sont pas applicables (WHITLOCK ET MCCAULEY, 1999). C'est notamment le cas lorsque l'on se place dans les conditions d'observation d'une maladie émergente. La démogénétique représente l'étude

conjointe de l'évolution démographique et génétique d'une population, et des interactions entre les deux. La prise en compte de ces interactions permet de s'affranchir de l'hypothèse d'équilibre démographique et de comprendre plus finement l'évolution des populations.

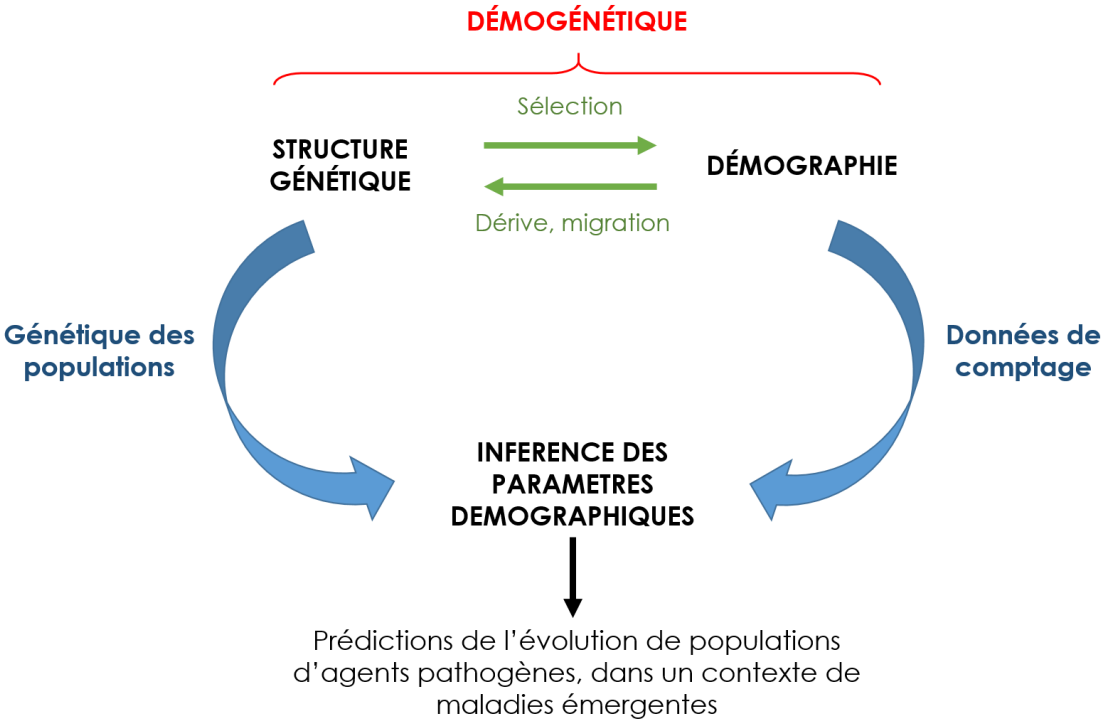


FIGURE 2 – Représentation schématique du lien entre la structure génétique et la démographie d'une population pour comprendre son évolution.

Les populations d'agents pathogènes responsables de maladies émergentes peuvent être soumises à de grandes variations démographiques, notamment des goulots d'étranglement et des événements de fondation répétés (GLADIEUX *et al.*, 2011). En zone tempérée, la plupart des espèces d'agents pathogènes présentent des cycles de vie saisonniers avec des alternance entre une phase épidémique et une phase de dormance (AGRIOS, 2005). Cette dernière phase correspond à une période de survie face à des conditions environnementales difficiles ou à l'absence d'hôte. La taille de la population de l'agent pathogène peut diminuer drastiquement pendant cette période, avant de réaugmenter lors de la saison épidémique. Ces variations extrêmes de tailles de populations rendent les agents pathogènes particulièrement vulnérables aux extinctions locales (LAINE ET BARRÈS, 2013) et à la dérive génétique. De plus, les populations d'agents pathogènes peuvent être soumises à une forte sélection, notamment lors des événements d'adaptation à l'hôte (GLADIEUX *et al.*, 2011). La sélection affecte la démographie de l'agent pathogène, à la fois *via* une mortalité accrue des individus les moins adaptés et *via* une augmentation de la reproduction des individus les plus adaptés (MCDONALD ET LINDE, 2002). Dans un contexte de maladie émergente, la compréhension de l'évolution des populations pathogènes nécessite donc avant tout de

comprendre le lien entre les variations démographiques et les variations génétiques de ces populations, grâce à la démogénétique.

De la modélisation à l'inférence

Un modèle est une représentation mathématique simplifiée de la réalité qui permet de poser des hypothèses et d'étudier les mécanismes en jeu dans des conditions contrôlées. La modélisation permet d'appréhender trois objectifs distincts : la description, la compréhension et la prévision. La modélisation présente un compromis entre réalisme et complexité : un modèle doit représenter de manière suffisamment réaliste le système étudié tout en le synthétisant le plus simplement possible (*i.e.* avec le moins de paramètres possible), deux facteurs qui sont souvent négativement corrélés. Le réalisme du modèle et la complexité associée dépendent donc des choix du modélisateur, en fonction du processus modélisé et du niveau d'interprétation recherché.

Les modèles appliqués aux systèmes biologiques permettent d'établir des attendus théoriques. La confrontation de ces attendus théoriques aux données réelles nous permet de tirer en retour des conclusions sur les processus biologiques. Si les hypothèses posées par un modèle sont suffisamment pertinentes par rapport au système biologique observé, la modélisation peut être utilisée au delà de la simple description et compréhension théorique pour aller jusqu'à une inférence des valeurs de certains paramètres du modèle, et donc du processus biologique sous-jacent. Arriver jusqu'à cette étape d'inférence mais rencontrer des problèmes d'estimation de certains paramètres peut également renseigner sur les limites d'un modèle et sur les pistes d'amélioration.

Parmi les méthodes d'inférence, deux seront utilisées dans le cadre de cette thèse :

1. Inférence par maximum de vraisemblance. Cette méthode d'inférence consiste à déterminer les paramètres qui maximisent la vraisemblance des données réelles, *i.e.* la probabilité d'observer ces données sous les hypothèses du modèle. La vraisemblance $L(\theta)$ correspond à la probabilité P d'observer un ensemble de données Y sous le modèle considéré, sachant les valeurs de paramètres exprimées ici par le vecteur θ :

$$L(\theta) = P(Y|\theta)$$

Pour un ensemble de n observations indépendantes $(y_1, y_2, ..., y_n)$, la vraisemblance $L(\theta)$ est le produit des vraisemblances de chaque observation :

$$L(\theta) = \prod_{i=1}^{n} P(y_i|\theta)$$

Cette méthode nécessite de pouvoir calculer la vraisemblance en fonction des paramètres du modèle, afin de déterminer le vecteur de paramètres θ^* qui la maximise. Pour ce faire, il est généralement nécessaire d'établir les lois statistiques qui régissent l'observation des données et qui permettent de faire le lien avec les valeurs des paramètres du modèle. Selon les modèles, la vraisemblance peut cependant être très coûteuse à évaluer et connaître son expression analytique n'est pas toujours possible.

2. Inférence par la méthode du Calcul Bayésien Approché, ou Approximate Bayesian Computation (ABC). L'inférence Bayésienne repose sur le théorème de Bayes qui décrit les probabilités conditionnelles des évènements A et B :

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

En appliquant ce théorème à un vecteur de paramètres θ à estimer en connaissant les données Y, nous obtenons :

$$P(\theta|Y) = \frac{P(Y|\theta)\ P(\theta)}{P(Y)}$$

Avec $P(\theta)$ la distribution *a priori* de θ, selon les hypothèses du modélisateur sur les valeurs de θ, et $P(\theta|Y)$ la distribution *a posteriori* de θ, selon les données observées Y. $P(Y|\theta)$ représente donc la vraisemblance d'observer les données Y selon le vecteur de paramètres θ. Cette vraisemblance peut être difficile à exprimer et à calculer numériquement, notamment lorsque le modèle est composé de nombreux paramètres (DANESH, 2021). Dans ce cas, une inférence par la méthode du Calcul Bayésien Approché ABC peut être utilisée. Elle repose sur l'utilisation de simulations pour établir la distribution *a posteriori* de θ. Celle-ci est déterminée en retenant la fraction des simulations qui produisent les résultats les plus proches des observations réelles. Cette méthode permet d'estimer les paramètres en se passant de la fonction de vraisemblance. Plus en détail, l'algorithme numérique repose sur les étapes suivantes : un plan de simulation est construit en tirant les paramètres de chaque simulation (ceux que l'on souhaite estimer à partir des données observées) dans des distributions *a priori*. Ces distributions *a priori* permettent de restreindre l'espace des paramètres à des valeurs biologiquement réalistes. Chaque combinaison de valeurs de paramètres du plan de simulation permet de réaliser une simulation qui calcule et renvoie un ensemble de statistiques résumées que l'on est aussi capable de calculer à partir des données observées. Pour un modèle de biologie évolutive, ces statistiques résumées peuvent représenter par exemple les indices de génétique des populations précédemment décrits (H_E, H_O, F_{IS}, F_{ST}, etc.). À partir de ces statistiques résumées (simulées et observées), l'analyse ABC permet ensuite de déterminer les distributions *a posteriori* des paramètres qui expliqueraient le mieux les données observées, d'après les étapes décrites en FIGURE 3 dans un cas simple.

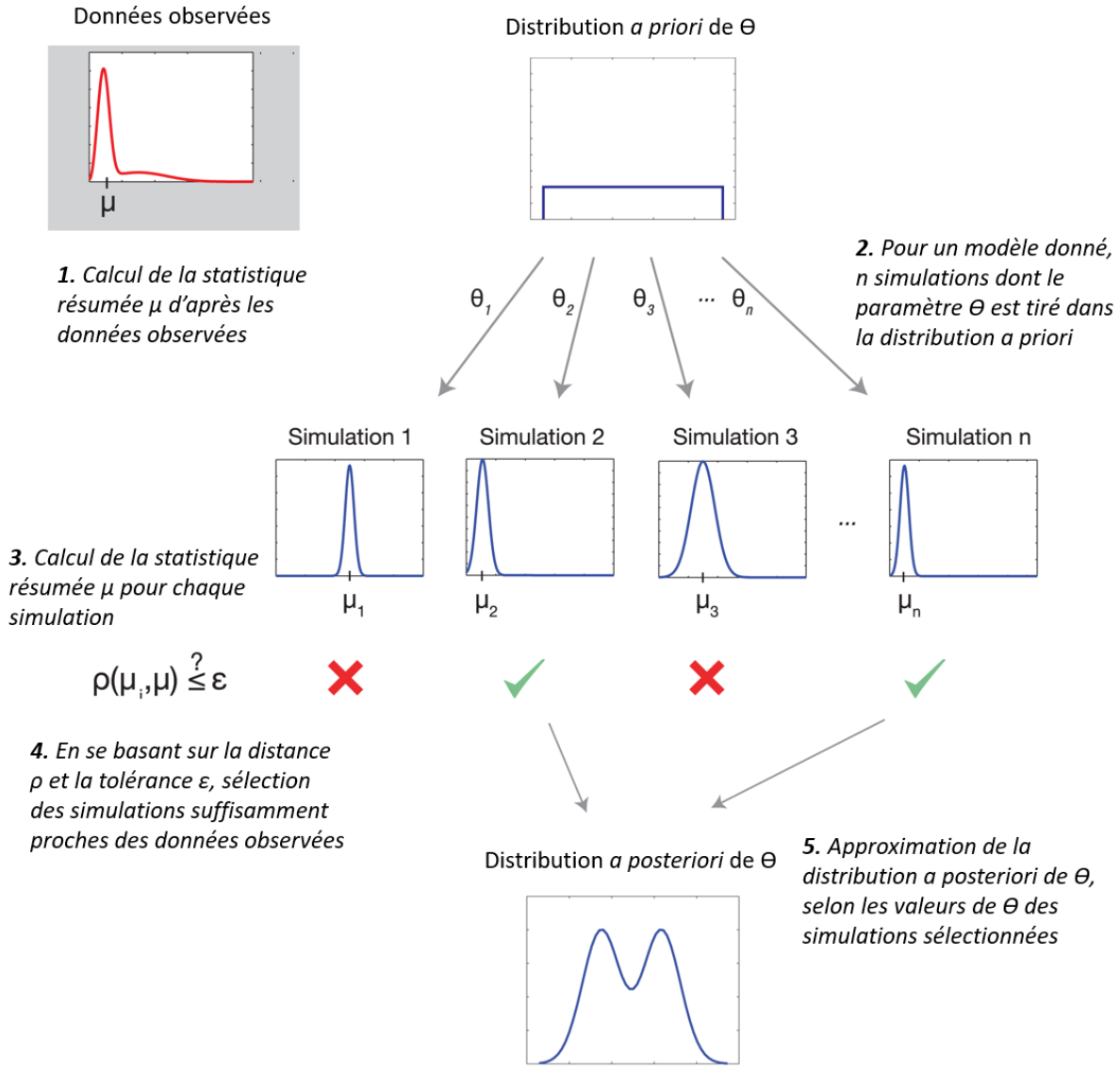


FIGURE 3 – Schéma de l'estimation d'un paramètre θ à partir de la statistique résumée μ par la méthode ABC (Adapté de SUNNAKER *et al.*, 2013).

Modèle d'étude : *Melampsora larici-populina*, agent de la rouille du peuplier

Melampsora larici-populina (Mlp) est l'agent pathogène majeur responsable de la maladie de la rouille du peuplier. C'est un champignon qui appartient à la division des Basidiomycètes, à la classe des Pucciniomycètes et à l'ordre des Pucciniales. Le genre *Melampsora* regroupe plus de 50 espèces dont 17 sont des agents pathogènes du peuplier (genre *Populus*, VIALLE *et al.*, 2011). Parmi les neuf espèces présentes sur le continent européen, Mlp est celle qui cause les dégâts les plus importants dans les peupleraies françaises (PINON ET FREY, 2005).

Il s'agit d'un organisme biotrophe, macrocyclique (avec la production de cinq types de spores) et hétéroïque (avec une alternance de reproductions sexuée et asexuées sur deux espèces d'hôte) (FIGURE 4). Ces caractéristiques en font, comme la plupart des agents des rouilles, un organisme au cycle de vie particulièrement complexe (DUPLESSIS *et al.*, 2021). Une reproduction sexuée a lieu au printemps sur l'hôte alternant (ou hôte écidien), le mélèze *Larix* sp. : les gamètes (basidiospores haploïdes) infectent le mélèze, générant ainsi les structures de reproduction sexuée (les spermogonies haploïdes qui produisent les spermaties haploïdes). L'étape de plasmogamie résulte ensuite du contact entre une spermatie haploïde et une spermogonie haploïde de signes sexuels compatibles. Des spores dicaryotiques (écidiospores) sont alors produites par les écidies (FIGURE 5a). La reproduction sexuée est donc directement suivie par un cycle de multiplication asexuée sur mélèze. Ces écidiospores infectent l'hôte télien, le peuplier (genre *Populus*), à la fin du printemps. Une dizaine de cycles de multiplication asexuée ont ensuite lieu sur le peuplier du printemps à l'automne *via* les urédies (FIGURE 5b). Lors de cette phase épidémique du cycle sur peuplier, Mlp est sous forme dicaryotique, ce qui revient à une forme diploïde. La caryogamie et le début de la méiose ont lieu à l'automne. Mlp survit à l'hiver sur les feuilles de peupliers sénescentes. Au début du printemps suivant, le cycle de vie reprend avec la fin de la méiose qui génère les basidiospores émises à partir des feuilles de peupliers dans la litière (HACQUARD *et al.*, 2011). La reproduction sexuée de Mlp est rendue obligatoire en climat tempéré par la chute annuelle des feuilles de peupliers (XHAARD *et al.*, 2011).

Les conséquences de l'infection sur le mélèze sont peu marquées (PERSOONS, 2015). A l'inverse, sans être létaux, les dommages sont bien plus marqués sur le peuplier. La multiplication des urédies lors de la phase épidémique entraîne une diminution des capacités photosynthétiques et une défoliation précoce du peuplier, ce qui provoque *in fine* une réduction de croissance (FIGURE 6) (GÉRARD *et al.*, 2006). Les grumes de peupliers sont récoltées 15 à 20 ans après leur plantation. Grâce à sa croissance rapide et ses utilisations variées (caisserie, bois énergie, contreplaqué industriel, etc.), le peuplier est la troisième essence de feuillus récoltée en France, derrière le chêne et le hêtre (IGN, 2022). En France, les peupleraies cultivées sont principalement présentes dans le Nord du territoire et dans les vallée alluviales (Loire, Garonne, Rhin). Jusqu'aux années 1980-1990, certains cultivars interaméricains, porteurs d'une résistance totale à Mlp, ont connu un grand succès commercial. Suite aux fortes attaques de rouilles sur ces cultivars, les plantations se sont diversifiées et les populiculteurs se sont tournés ces dernières décennies vers les cultivars euraméricains présents sur le marché (FCBA, 2020). Malgré la diversification des cultivars plantés, chaque peupeuraie cultivée reste cependant monospécifique et monoclonale. En plus de ces peupliers cultivés, des peupliers sauvages (*Populus nigra*) sont présents sur l'ensemble sur territoire français. Mlp infecte indifféremment les peupliers cultivés et sauvages, mais les conséquences de cette infection sont bien plus sévères sur les peupliers

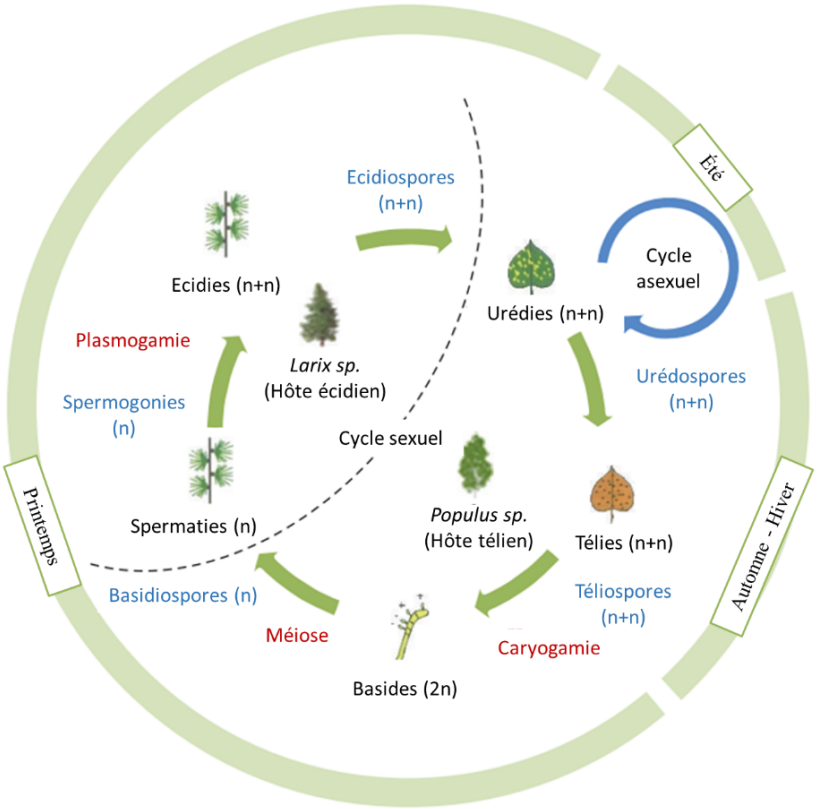


FIGURE 4 – Cycle de vie hétéroïque de l'agent pathogène *Melampsora larici-populina* (d'après HACQUARD *et al.*, 2011).

cultivés du fait notamment de leurs conditions de culture (FREY *et al.*, 2005).

Du fait des contournements de résistances observés, Mlp est un agent pathogène responsable d'une maladie ré-émergente, *i.e.* une maladie déjà connue qui réapparaît sur un territoire donné. Cet organisme modèle a déjà été longuement étudié dans l'équipe d'accueil (XHAARD *et al.*, 2009, 2011; XHAARD, 2011; XHAARD *et al.*, 2012; BECHELER *et al.*, 2016; PERSOONS, 2015; PERSOONS *et al.*, 2017; MAUPETIT, 2019; LOUET *et al.*, 2021), et de nombreuses données empiriques sont disponibles, avec une collection d'individus de Mlp dont la plupart sont génotypés. Cet agent pathogène représente donc un très bon organisme modèle pour étudier l'évolution rapide d'un agent pathogène au cycle de vie complexe dans le cadre d'une maladie émergente.



(a) (b)

FIGURE 5 – Photographies (a) des écidies de Mlp sur une aiguille de mélèze (*Source : P. Frey*), et (b) des urédies de Mlp sur une feuille de peuplier (*Source : B. Petre*).



(a) (b) (c)

FIGURE 6 – Photographies des différents impacts de Mlp sur les peupleraies cultivées françaises. (a) Peupleraie saine ; (b) peupleraie infectée ; (c) peupleraie fortement infectée, et dont une partie des peupliers sont morts (*Source : P. Frey*).

Problématiques et objectifs

Il est difficile d'étudier l'évolution des populations lorsqu'elles présentent de fortes variations démographiques, comme c'est le cas des agents pathogènes en contexte de maladie émergente. Dans cette thèse, je souhaite mettre en avant le lien entre la démographie et l'évolution génétique d'une population d'agents pathogènes responsables de maladies émergentes ou ré-émergentes. Les objectifs sont de comprendre les conditions d'observations de telles structures génétiques, d'inférer les paramètres démographiques de la population, et de prédire son évolution spatio-temporelle.

Les informations et les données disponibles sur l'agent pathogène Mlp en font un très bon organisme modèle pour étudier les différentes conditions d'observation d'une maladie émergente et les structures génétiques et démographiques associées. Dans cette thèse,

j'utilise la démogénétique et l'apport conjoint de la modélisation et des données empiriques pour étudier la maladie ré-émergente de la rouille du peuplier. Pour cela, je me place volontairement dans deux situations où l'hypothèse de l'équilibre démographique n'est pas respectée, mais que l'on peut modéliser :

1. **Un contournement de résistance** représentant une maladie ré-émergente réelle *via* l'adaptation de l'agent pathogène à son hôte. Dans ce système, la composante **temporelle** est déterminante.
2. **Plusieurs expansions clonales annuelles récurrentes le long d'une même vallée**, dont la connaissance de la population source en fait un très bon système pour l'étude d'une maladie ré-émergente géographique, pour laquelle la composante **spatiale** est déterminante.

L'étude de chaque situation écologique forme une partie distincte de cette thèse. Chacune de ces deux parties regroupe une association théorique et empirique *via* la comparaison des structures génétiques observées et prédites par les modèles construits, jusqu'à l'inférence de certaines valeurs de paramètres. Ces deux parties sont organisées de façon symétrique avec un chapitre introductif détaillant les spécificités de ces cadres écologiques et évolutifs et les hypothèses de modélisation qui peuvent en découler, suivi de plusieurs chapitres regroupant les résultats obtenus. Je conclus ce travail par une discussion générale des effets de la démographie sur la structure de populations hors équilibre dans le cadre d'une maladie émergente.

Première partie

SUIVI TEMPOREL D'UNE MALADIE RÉ-ÉMERGENTE : LE CAS D'UN CONTOURNEMENT DE RÉSISTANCE

Objectifs

1. Déterminer les conditions d'existence d'un contournement de résistance par la modélisation, et notamment l'influence du niveau de ploïdie et du cycle de vie de l'agent pathogène.
2. Déterminer l'interaction entre les variations démographiques et la structure génétique neutre résultant d'un évènement de contournement de résistance.
3. Inférer le scénario de contournement ainsi que les paramètres écologiques et démographiques ayant régi le contournement de la résistance RMlp7 par l'analyse conjointe des données empiriques et du modèle.

Chapitre 1 Introduction **19**
1.1 Les contournements de résistances . . . 21
1.2 Article 1. Improving sustainable crop protection using population genetics concepts . . . 26
1.3 Cas d'étude : le contournement de la résistance RMlp7 par Mlp . . . 61
1.4 Objectifs . . . 64

Chapitre 2 Étude épidemio-évolutive d'un contournement de résistance **67**
2.1 Article 2. Impact of ploidy and pathogen life cycle on resistance durability . . . 69
2.2 Analyse déterministe par résolution analytique . . . 101
2.2.1 Enjeux de la résolution déterministe . . . 101
2.2.2 Description du modèle . . . 101
2.2.3 Les différents équilibres et leur stabilité . . . 104
2.2.3.1 Extinction de la population . . . 105
2.2.3.2 Disparition des individus hétérozygotes Aa . . . 107
2.2.3.3 Disparition des individus homozygotes AA . . . 109
2.2.3.4 Fixation de l'allèle A . . . 111
2.2.3.5 Fixation de l'allèle a . . . 114
2.2.3.6 Survie des individus hétérozygotes Aa uniquement . . . 115

2.2.3.7 Conditions restantes 116
2.2.4 Conclusion 119
2.3 Conclusion en regard du contournement de la résistance RMlp7 par Mlp . . . 119

Chapitre 3 Etude démogénétique d'un contournement de résistance 123
3.1 Article 3. Interplay between demography and selection: a forward model illuminates the temporal genetic signatures of rapid adaptation in plant pathogens 125
3.2 Conclusion en regard du contournement de la résistance RMlp7 par Mlp . . . 157

Chapitre 4 Inférence d'un scénario de contournement 161
4.1 Article 4. Approximate Bayesian Computation applied to time series of population genetic data disentangles rapid genetic changes and demographic variations 163
4.2 Inférence du scénario de contournement de la résistance Rmlp7 193

Conclusion de la partie I 195

Chapitre 1

Introduction

1.1 Les contournements de résistances

Pour lutter contre les agents pathogènes en agriculture tout en limitant l'utilisation des pesticides, la sélection de résistances génétiques qualitatives chez les plantes est largement utilisée. Une résistance qualitative est basée sur l'interaction gène-pour-gène (FLOR, 1971) et repose sur un gène majeur de résistance pour empêcher toute infection de la plante hôte par l'agent pathogène. Dans le cas le plus simple d'une réponse spécifique, le résultat de l'infection est déterminé par l'interaction entre un locus de la plante (le gène de résistance) et un locus du pathogène (le gène d'avirulence). Le mécanisme de réponse est alors de type "tout ou rien" avec le blocage ou non de l'infection par la plante. Les résistances qualitatives reposent souvent sur la reconnaissance d'une molécule spécifique produite par l'agent pathogène (une protéine effectrice par exemple) par un récepteur immunitaire de la plante (LO PRESTI *et al.*, 2015). Cependant les agents pathogènes peuvent évoluer rapidement et s'adapter à leur hôte. Lorsqu'une variété cultivée possède de bonnes propriétés agronomiques, notamment *via* sa capacité à résister aux infections, elle est souvent largement plantée dans les paysages agricoles. Cela augmente la pression de sélection unidirectionnelle exercée sur les populations d'agents pathogènes. L'apparition d'un individu virulent dans la population d'agents pathogènes lui donne alors un avantage sélectif suffisant pour que l'allèle responsable de la virulence augmente rapidement en proportion dans la population, et que les individus virulents envahissent les hôtes résistants. On parle alors de contournement de résistance (GOYEAU *et al.*, 2021).
Un tel contournement de résistance se déroule en trois étapes (FIGURE 1.1) (GOYEAU *et al.*, 2021) :

1. **L'apparition** d'un individu virulent. L'allèle virulent peut apparaître par mutation, ou être déjà présent dans la population sous forme de variabilité génétique préexistante. Pour une espèce haploïde, l'apparition de l'allèle virulent au sein d'un individu se traduit *de facto* par l'apparition d'un individu virulent (FIGURE 1.2a). Lorsque le produit du gène d'avirulence est une molécule spécifique comme une protéine effectrice, le pathogène n'est virulent que si ce produit n'est pas détecté par le produit du gène de résistance correspondant chez l'hôte (STUKENBROCK ET MCDONALD, 2009). Par conséquent, un agent pathogène diploïde n'est virulent que si aucun des produits des deux allèles n'est détecté par l'hôte. En d'autres termes, dans l'interaction classique gène pour gène l'allèle virulent est récessif (FIGURE 1.2b) (THRALL *et al.*, 2016). Un individu virulent diploïde (*i.e.* homozygote pour l'allèle virulent) peut apparaître suite à l'étape de recombinaison.
2. **La multiplication** responsable de l'augmentation en proportion des individus virulents dans la population. Deux forces sélectives sont impliquées dans cette multiplication : la sélection et la dérive génétique (GOYEAU *et al.*, 2021). L'individu virulent pos-

sède un avantage sélectif car il est capable d'infecter un hôte résistant. Au moment de son apparition, cet avantage sélectif est d'autant plus élevé qu'il est le seul individu virulent dans la population. Il pourra donc plus facilement accéder aux ressources produites par l'hôte résistant et se reproduire, entraînant ainsi une augmentation de la proportion de l'allèle virulent dans la population (MÖLLER ET STUKENBROCK, 2017). Cependant l'allèle virulent ne confère un avantage sélectif à son porteur que lorsqu'il se trouve sur un hôte résistant. Sur un hôte sensible, l'allèle virulent n'est pas nécessairement soumis à la sélection mais reste soumis à la dérive génétique. La force de la dérive génétique dépend alors de la taille de la population. Cette étape de dérive génétique précédant la sélection peut entraîner un délai entre l'apparition d'un individu virulent et le contournement de résistance, voire même une disparition de l'allèle virulent de la population. C'est une source importante de stochasticité.

3. **La dispersion** des individus virulents sur de nombreux hôtes résistants. Les individus virulents, en grande proportion dans la population, peuvent se disperser *via* la migration sur les hôtes sensibles et résistants et envahir rapidement des parcelles agricoles composées de variétés résistantes. C'est l'étape de crise épidémique qui mène à la détection par l'homme du contournement de la résistance.

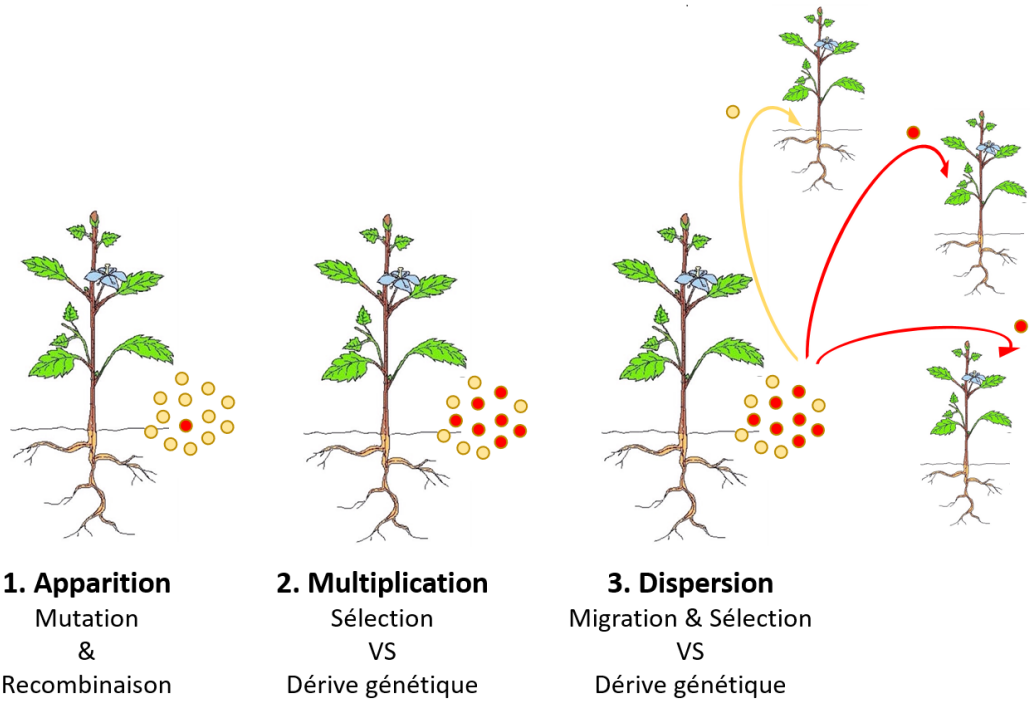
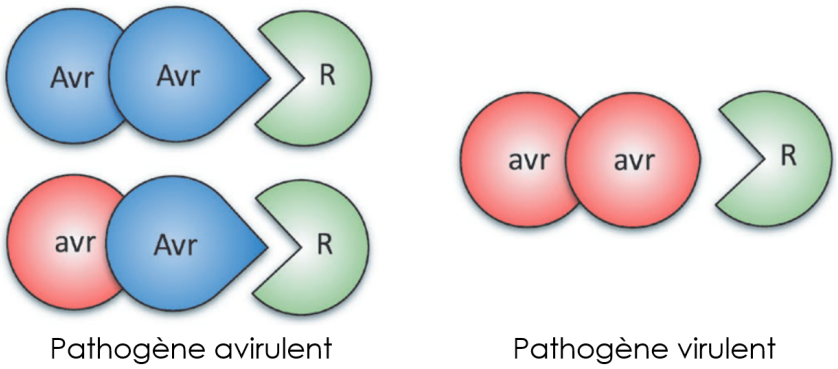


FIGURE 1.1 – Schéma des trois étapes d'un contournement de résistance et des forces évolutives impliquées. Les ronds représentent les agents pathogènes : en jaune les individus avirulents capable d'infecter les hôtes sensibles uniquement et en rouge les individus virulents capables d'infecter les hôtes sensibles et résistants (d'après GÓMEZ *et al.*, 2009 et GOYEAU *et al.*, 2021).



(a) Le cas d'un agent pathogène haploïde



(b) Le cas d'un agent pathogène diploïde

FIGURE 1.2 – Schéma de la différence dans le déterminisme moléculaire de la résistance entre les pathogènes (a) haploïdes et (b) diploïdes (d'après LOUET, 2021). Avr et avr représentent les produits des allèles avirulent et virulent de l'agent pathogène, respectivement. R représente un récepteur immunitaire de l'hôte résistant, capable de reconnaître uniquement le produit Avr de l'allèle avirulent de l'agent pathogène. En cas de détection d'au moins un produit Avr par le récepteur immunitaire de l'hôte, l'infection est stoppée et le pathogène est avirulent. En cas de non détection des produits avr par le récepteur de l'hôte, l'infection a lieu et le pathogène est virulent.

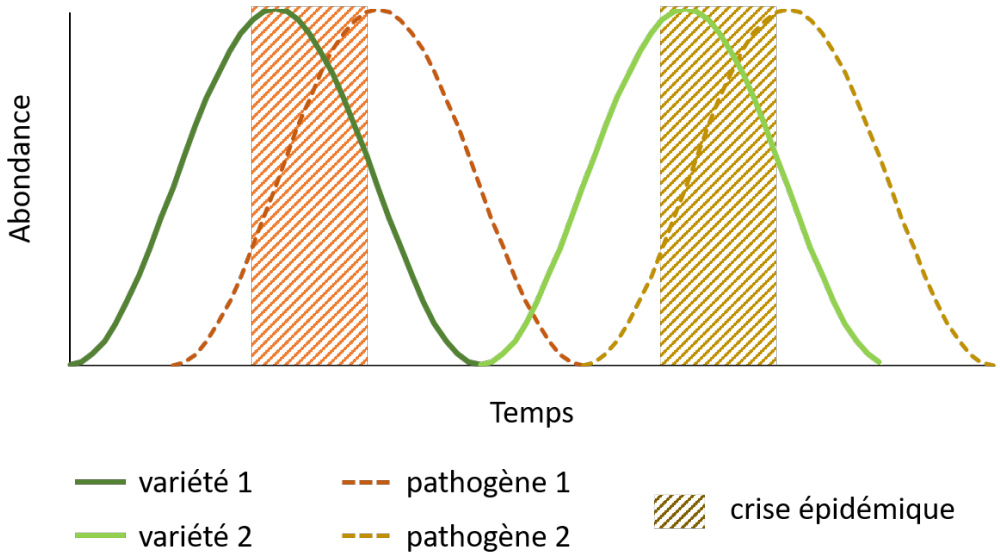
La détection d'un contournement de résistance en paysage agricole entraîne souvent le retrait de la variété résistante du marché, puis le déploiement d'une nouvelle résistance. Ces successions de déploiements de résistances et de leurs contournements sont appelés cycles boom-and-bust (FIGURE 1.3a). Ces cycles sont principalement documentés chez les grandes cultures (le blé, l'orge, le colza, le riz, ou l'avoine, FIGURE 1.3b)

De tels contournements ont des conséquences économiques importantes et menacent également la sécurité alimentaire lorsque la plante hôte concernée est cultivée pour sa consommation. La souche Ug99 de l'agent pathogène de la rouille noire *Puccinia graminis* f. sp. *tritici* est l'un des exemples les plus caractéristiques de menace de la sécurité alimentaire. Cette souche a été détectée pour la première fois en Ouganda en 1998 et possède une virulence qui contourne un grand nombre de gènes de résistance présents dans les variétés de blé cultivé *Triticum aestivum* (SINGH *et al.*, 2015). Cet agent pathogène menace actuellement les cultures de blé de l'ensemble de l'Afrique de l'Est.

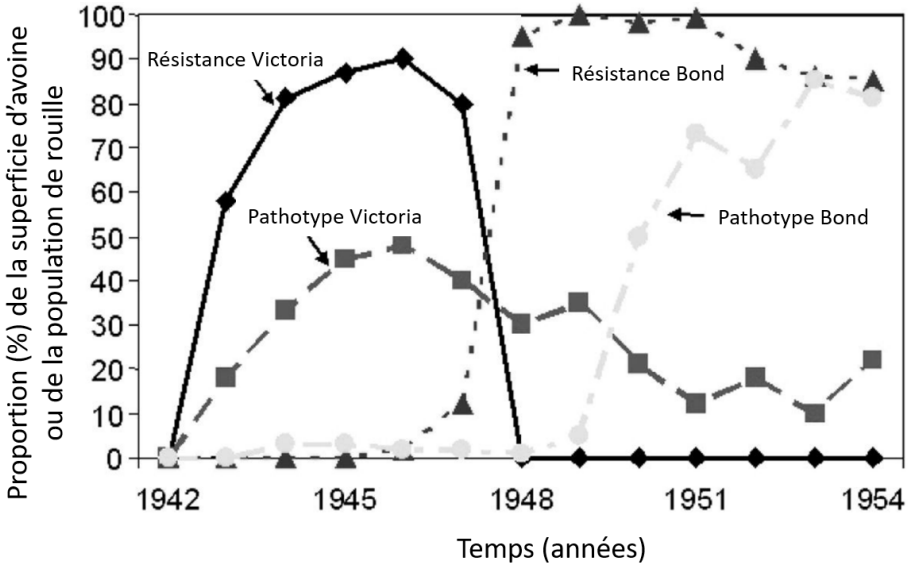
Les contournements de résistances peuvent être plus ou moins rapides selon le pathosystème et les conditions environnementales (VAN DEN BOSCH ET GILLIGAN, 2003). Cela

soulève la question de la durabilité d'une résistance (JOHNSON, 1979), que l'on peut définir comme le temps écoulé entre le déploiement d'une résistance et son contournement par un agent pathogène, *i.e.* lorsque la proportion d'individus virulents dans la population atteint un seuil donné. La définition de la durabilité d'une résistance peut varier d'une étude à l'autre selon le seuil considéré (PIETRAVALLE *et al.*, 2006; BROWN, 2015; CAROLAN *et al.*, 2017; LOF ET VAN DER WERF, 2017; PACILLY *et al.*, 2018; RIMBAUD *et al.*, 2021). Une agriculture durable doit réussir à réduire les épidémies et contrôler sur le long terme l'évolution des populations d'agents pathogènes (RIMBAUD *et al.*, 2021).

Un autre type de résistances, les résistances quantitatives, sont sélectionnées pour être plus tolérantes aux agents pathogènes, mais sans conférer d'immunité complète. Ces résistances permettent ainsi de limiter l'impact des agents pathogènes sur l'hôte tout en évitant les crises épidémiques majeures telles que celles observées lors des contournements de résistances qualitatives. Ces résistances quantitatives sont supposées plus durables (COWGER ET BROWN, 2019) car elles réduisent la pression de sélection exercée sur l'agent pathogène. Elles sont cependant plus difficiles à sélectionner et peuvent entraîner une adaptation des agents pathogènes vers une agressivité plus élevée (COWGER ET MUNDT, 2002; PARIAUD *et al.*, 2009) et des dégâts plus importants sur l'hôte (GANDON ET MICHALAKIS, 2000).



(a) Schéma des cycles boom-and-bust



(b) Exemple des contournements de résistance successifs par la rouille de l'avoine

FIGURE 1.3 – Les cycles boom-and-bust. (a) Représentation schématique des cycles boom-and-bust, avec l'augmentation des surfaces cultivées des variétés résistantes (boom) puis la régression de ces surfaces au fur et à mesure des contournements (bust) (d'après GOYEAU *et al.*, 2021). (b) Exemple de cycles boom-and-bust documentés chez la rouille de l'avoine dans l'Iowa, avec les contournements successifs de deux résistances portées par les variétés d'avoine Victoria et Bond (adapté de MCDONALD, 2004 d'après les données de BROWNING ET FREY, 1969).

1.2 Article 1. Improving sustainable crop protection using population genetics concepts

Le nombre d'études s'intéressant et mentionnant la durabilité des résistances chez les plantes est très élevé, et augmente régulièrement depuis les années 1990. La question de la durabilité des résistances peut être étudiée à des échelles très différentes : depuis les mécanismes moléculaires menant à la résistance, jusqu'aux conséquences épidémiologiques des contournements. Une analyse bibliographique du corpus de recherche étudiant la durabilité des résistances chez les plantes nous a permis de mettre en évidence un fossé entre 3 domaines de recherche : 1) le domaine de la sélection variétale, 2) le domaine de l'étude des interactions moléculaires plantes-pathogènes et 3) le domaine de l'épidémiologie et de l'évolution. Nous pensons que ce fossé freine considérablement les avancées de la recherche sur la durabilité des résistances. Pour remédier à cette situation, nous proposons d'utiliser les concepts de la génétique des populations pour établir des liens plus solides entre ces trois disciplines.

Cette analyse est présentée ici sous la forme d'un article d'opinion intitulé « Improving sustainable crop protection using population genetics concepts », qui a été accepté pour publication en juin 2022 dans la revue MOLECULAR ECOLOGY.

Improving sustainable crop protection using population genetics concepts

Méline SAUBIN[1], Clémentine LOUET[1,2], Lydia BOUSSET[3], Frédéric FABRE[4], Pascal FREY[1], Isabelle FUDAL[2], Frédéric GROGNARD[5], Frédéric HAMELIN[3], Ludovic MAILLERET[5,6], Solenn STOECKEL[3], Suzanne TOUZEAU[5], Benjamin PETRE[1], Fabien HALKETT[1*]

[1] Université de Lorraine, INRAE, IAM, Nancy, France
[2] Université Paris Saclay, INRAE, BIOGER, F-78850 Thiverval-Grignon, France
[3] INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, Le Rheu, France
[4] INRAE, Bordeaux Sciences Agro, SAVE, F-33882 Villenave d'Ornon, France
[5] Université Côte d'Azur, Inria, INRAE, CNRS, Sorbonne Université, Biocore team, Sophia Antipolis, France
[6] Université Côte d'Azur, INRAE, CNRS, ISA, Sophia Antipolis, France
***** Corresponding author (fabien.halkett@inrae.fr)

Published in Molecular Ecology, DOI: https://doi.org/10.1111/mec.16634

Abstract

Growing genetically resistant plants allows pathogen populations to be controlled and reduces the use of pesticides. However, pathogens can quickly overcome such resistance. In this context, how can we achieve sustainable crop protection? This crucial question has remained largely unanswered despite decades of intense debate and research effort. In this study, we used a bibliographic analysis to show that the research field of resistance durability has evolved into three subfields: (i) *'plant breeding'* (generating new genetic material), (ii) *'molecular interactions'* (exploring the molecular dialogue governing plant–pathogen interactions) and (iii) *'epidemiology and evolution'* (explaining and forecasting of pathogen population dynamics resulting from selection pressure(s) exerted by resistant plants). We argue that this triple split of the field impedes integrated research progress and ultimately compromises the sustainable management of genetic resistance. After identifying a gap among the three subfields, we argue that the theoretical framework of population genetics could bridge this gap. Indeed, population genetics formally explains the evolution of all heritable traits, and allows genetic changes to be tracked along with variation in population dynamics. This provides an integrated view of pathogen adaptation, in particular via evolutionary–epidemiological feedbacks. In this Opinion Note, we detail examples illustrating how such a framework can better inform best practices for developing and managing genetically resistant cultivars.

Keywords

Plant immunity, Resistance durability, Citation network, Transient dynamics, Genetic drift, Host-pathogen coevolution

1. Genetic resistance durability

Growing genetically resistant plants helps regulate pathogen populations while considerably reducing pesticide use. Genetic resistances are used extensively in crops to confer partial or complete host immunity. Whereas some resistance genes, especially those relying on pathogen recognition, can be defeated quickly by pathogens, leading to a loss of efficiency (Brown, 2015; REX Consortium, 2016), others can remain undefeated for decades (Cowger & Brown 2019; Piffanelli 2004). The two crucial questions are thus: (1) why is there such variation in the durability of resistances, and (2) how can resistance durability be enhanced? This Opinion Note embraces an evolutionary view of plant–pathogen interactions for a better understanding of pathogen adaptation in the face of plant resistances, highlighting the role of theoretical population genetics in achieving sustainable management of plant diseases.

2. The plant resistance durability literature comprises three distinct subfields

This Opinion Note was prompted by an analysis of the literature on plant resistance durability (Annex, Fig. S1–S8) that revealed a large split among studies. A keyword co-occurrence analysis distinguished three subfields of research (Fig. 1), hereafter referred to as '*plant breeding', 'molecular interactions'* and *'epidemiology and evolution'*, each of which addresses a different scale of pathogen adaptation to plant resistances. These subfields share the potential to ultimately improve plant health and increase the durability of crop protection, but they tackle these issues in very different ways (Box 1).

The observed split may have originated from at least two causes of divergence. First, these subfields focus on different stages, upstream or downstream of resistance durability, ranging from the identification of new resistances to resistances being defeated, in particular because of their wide deployment in agroecosystems. Typically, the selection of new plant resistances is the goal of the *'plant breeding'* subfield. Second, for studies that focus on already identified plant resistances, the split originates from a difference in viewpoint. *'Molecular interactions'* studies focus on the mechanisms of resistance, whereas *'epidemiology and evolution'* studies aim to define better management policies. This amounts to a classical partition between proximal and distal points of view that examine the causes and consequences of a phenomenon, respectively. More recently, we have observed some changes in paradigms, for example, the search for the Holy Grail of durable resistance mechanisms versus the sustainable management of not-necessarily durable resistances (Rimbaud *et al*., 2021). As such, the origins of this gap remain multifactorial, and further characterisation warrants a dedicated epistemological study, which is beyond the scope of this Opinion Note.

An analogous split had already been documented in terms of resistance to pesticides (REX Consortium, 2010). Insights from multiple fields can leap forward scientific progress. Isolated contributions preclude the emergence of an integrated knowledge. Regarding the question of resistance durability, the split ultimately compromises the sustainable management of genetic resistances. Here, we suggest it might be possible to combine these three subfields of research to improve plant health and increase the durability of crop protection.

Box 1: Three-way split among plant resistance durability studies

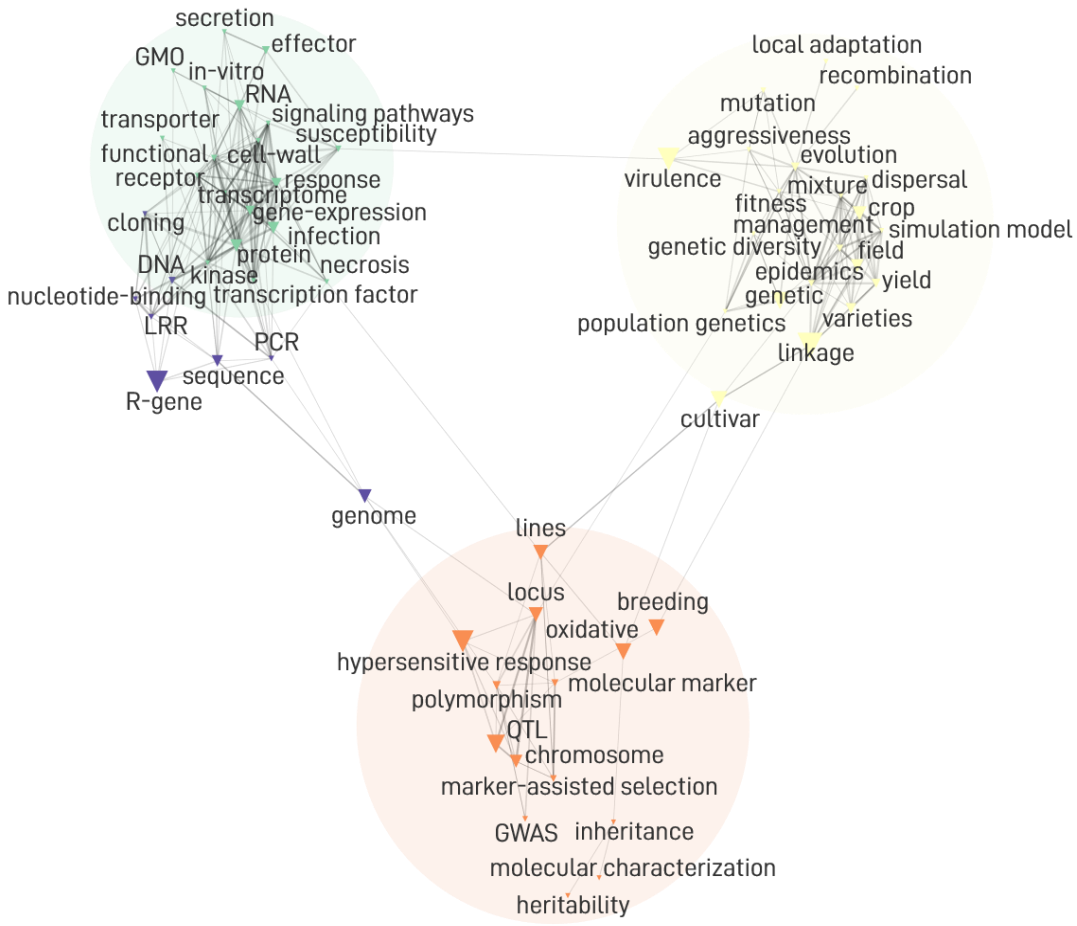


Fig. 1 Network of the co-occurrence of article keywords. Drawings were prepared using the Network Mapping tool on CorTexT Manager platform (https://www.cortext.net/), based on the refined list of 62 keywords and 1783 articles from the *'plant breeding', 'molecular interactions'*, and *'epidemiology and evolution'* clusters (See Annex for details, Table S1). The proximity threshold was fixed at 0.44, and all other parameters were set to default values. Nodes represent keywords; lines between nodes represent links that measure the mutual occurrence of the keywords among articles. The closer the nodes, the tighter the association. An interactive version of this figure is available at: https://documents.cortext.net/lib/mapexplorer/explorerjs.html?file=https://assets.cortext.net/docs/438fce7b216d23e2d2316cab168d14e9).

'Plant breeding' studies focus on the creation of new plant varieties with, among other goals, intrinsically durable resistances. Historically, plant breeding relied on traditional phenotypic selection, and later integrated molecular markers by applying marker-assisted selection based on the identification of quantitative trait loci (Ribaut & Hoisington, 1998). The denser the markers, the more accurate the marker–trait association. More recently, the advent of next-generation sequencing techniques, and the resulting drop in whole-genome sequencing costs, led to the appearance of genomic selection, which uses the highest density of molecular markers to predict the performance of candidate genotypes for breeding. This method reduces the costs and time required for variety development (Crossa *et al.*, 2017). This shift in methodology also enables screening of an ever-growing source of new genetic resistances (from natural genetic variability to mutant lines). Notably, mutagenesis experiments have already produced very efficient and durable resistances, for example the *mlo-9* allele in barley (Jørgensen, 1992).

Because it confers complete host immunity, the gene-for-gene model (Flor, 1971) has attracted a great deal of interest and breeding efforts to select for varieties carrying qualitative resistances. According to the gene-for-gene model, resistance often relies on the recognition of a specific pathogen molecule (e.g. an effector protein, called an avirulence protein when recognised by a cognate resistance protein) by a plant receptor (Jones & Dangl, 2006). In such cases, pathogens can evolve with the loss or modification of the recognized molecule and thus escape host recognition, thereby restoring infectivity (Hartmann *et al*., 2017). Nevertheless, overcoming genetic resistance is not systematic (García-Arenal *et al.*, 2003; Nilusmas *et al*., 2020). The speed at which a given resistance is overcome is one of the main determinants of the duration of its use. Quantitative resistances have regained attention in recent years as a longer resistance durability can be expected (Cowger & Brown, 2019). Indeed, by conferring only a partial level of resistance to the plant, they also reduce the selection pressure on pathogen populations. Quantitative resistances can also increase the durability of qualitative resistances when combined within the same variety (Brun *et al.*, 2010; Fournet *et al.*, 2013; Palloix *et al.*, 2009; Laloi *et al.*, 2017). However, in a theoretical study, Gandon & Michalakis (2000) predicted that using quantitative resistance would select for increased levels of host damage. This agrees with empirical studies in which quantitative resistances have been shown to select for higher aggressiveness (Cowger & Mundt, 2002; Pariaud *et al.*, 2009). Finally, resistances that are not based on pathogen recognition can be far more durable. For example, barley plants carrying the recessive loss-of-function alleles (*mlo*) preventing hyphal penetration (Kusch & Panstruga, 2017) have so far proven resistant against all known isolates of the powdery mildew fungus *Blumeria graminis* f. sp. *hordei* (Lyngkjær *et al.*, 2000; Piffanelli *et al.*, 2004; Brown, 2015).

'Molecular interactions' studies examine how plant pathogens promote infection by altering plant defence reactions and plant cellular processes, and how plants recognize pathogens and trigger immunity. Technological innovations (i.e. genomics, bioinformatics) and strong conceptual models (i.e. recognition-based immunity, or the zig-zag model; Jones & Dangl, 2006) guide this subfield. In particular, cloning of resistance genes and associated avirulence genes (see, e.g., Dangl *et al.*, 2013) has led to three important discoveries that could help manage resistances and predict their durability:

- (1) The switch to virulence may have a fitness cost. Inactivation of avirulence genes or generation of near isogenic strains of fungi differing by the presence/absence of one avirulence gene allow measurement of potential fitness costs associated with a switch to virulence (Huang *et al.*, 2006). These fitness costs differ depending on the avirulence gene (Nováková *et al.*, 2016), and have major consequences on resistance durability and, therefore, on management strategies (Fabre *et al.*, 2012; Leach *et al.*, 2001).

- (2) The switch to virulence often results from alteration of the avirulence gene. Several molecular mechanisms are involved in the switch to virulence, with direct or indirect alteration of the avirulence gene. Depending on the avirulence effectors and their importance for pathogenesis, we observe either drastic mechanisms of gene inactivation (deletions, accumulation of mutations) or mechanisms that make it possible to escape recognition while maintaining the effector function (point mutations, intervention of another effector that masks recognition of the first) (Gómez *et al.*, 2009; Rouxel & Balesdent, 2017; Sánchez-Vallet *et al.*, 2018; Louet *et al.*, 2022). The variability of these mechanisms of transition to virulence can have important evolutionary consequences. Knowledge of such mechanisms could give clues as to the importance of the targeted effectors for the pathogen, and therefore the potential durability of the corresponding resistances (Janzac *et al.* 2009).

- (3) The gene-for-gene model fails to describe 'unconventional' molecular interactions. While an increasing number of avirulence genes have been identified in pathogens infecting agronomically important crops, 'unconventional' gene-for-gene interactions have also been reported. The recognition of one avirulence gene by several resistance genes, or epistatic interactions between avirulence genes, are typical examples of such interactions. All these interactions have different consequences for the switch from avirulence to virulence and, hence, on resistance gene management (for a review, see Petit-Houdenot & Fudal 2017).

'Epidemiology and evolution' studies decipher the evolutionary causes and consequences of pathogen adaptation at the population scale, employing both empirical and modelling approaches. To improve understanding of resistance durability, these studies first aim to identify the key drivers of durability. Several strategies differing in the type of resistance used and their deployment in time and space can be distinguished (Brown, 2015; REX Consortium, 2016). While knowledge in the *'molecular interactions'* field is spurred by technological advances brought about in the -omics era, *'epidemiology and evolution'* studies are driven mostly by field observations and include a large corpus of modelling works (Brown, 2015; Rimbaud *et al.* 2021). A major recent change in *'epidemiology and evolution'* studies is the increasing interest in quantitative resistances, as evidenced in our bibliographic analysis (Annex). Historically, epidemiological modelling studies started with the simple situation of plants bearing a single qualitative resistance. Recent studies encompass more complexity, in line with the observation of the overcoming of qualitative resistance alleles following their deployment. This prompted the field to increase investigation of the contribution of multiple and quantitative resistances on resistance durability (Lo Iacono *et al.*, 2012; Rimbaud *et al.*, 2018b; Zhan *et al.*, 2015). Some studies compare two resistance types with different deployment strategies, like resistance mixtures and pyramiding (REX Consortium, 2016; Rimbaud *et al.* 2018a; Djidjou-Demasse *et al.*, 2017). During the last two decades, there has been a trend to widen the quest for intrinsic durable resistance—resulting from the properties of resistance genes only—towards including the consideration of durable strategies of resistance management as an emerging property of clever agricultural practices (Rimbaud *et al.*, 2021). A key notion for increasing resistance durability would be to break the continuity in the selection pressure imposed by modern agricultural practices (Brown, 2015; Zhan *et al.*, 2015). Overall, this research subfield has led to three main findings so far:

- (1) It is crucial to take pathogen population biology into account. In particular, McDonald & Linde (2002) reviewed the typology of risk of overcoming resistance dependent on pathogen life cycle, finding that pathogens with large population sizes, mixed reproductive modes (i.e. both sexual and asexual reproduction) and massive, long-distance dispersal, have the highest evolutionary potential.

- (2) There is no absolute rule about the best deployment strategy to achieve resistance durability. All factors being equal, gene pyramiding will perform better in the case of *de novo* mutations, whereas other strategies (e.g. crop rotations and mixtures) are better suited to impeding pathogen evolution from standing genetic variation (Djidjou-Demasse *et al.*, 2017; REX Consortium, 2016; Rimbaud *et al.*, 2018a).

- (3) Chosen management strategies may differ depending on the time scale considered: short-term disease control or long-term mitigation of pathogen evolution (Brown, 2015; Papaïx *et al.*, 2018).

3. Population genetics provides a conceptual link across research subfields

In this Opinion Note, we advocate that population genetics provides us with the dedicated theory to merge the efforts of all three subfields (Box 2). Evolution results from changes in the frequencies of gene states (alleles), or any other heritable character. It results from the combined action of evolutionary forces, namely selection, migration, mutation, and drift. Population genetics is one of the rare established and efficient theoretical paradigms of biology, and relies strongly on mathematical modelling (Wakeley, 2005). Recent technological advances in investigations of molecular variations and genotyping have important ramifications for experimental sciences. Here, we build our proposal by considering that population genetics— more precisely, demogenetics—provides the concepts for knitting communities together by its inherent ability to integrate biological features and species interactions over time and space. This view differs from the current and traditional use of population genetics mostly as a tool to describe the genetic structure of populations within a species (see Annex for details). In the following, our proposal tackles how population genetics can help better explore plant–pathogen interactions towards the rational breeding and clever management of plant genetic resistances.

4. Knitting communities together with population genetics to improve plant health durability.

Our proposal is based on our analysis of the specificities of each of the three subfields, and identification of key research topics that would aid better subfield integration.

4.1 Consider evolutionary trajectories to highlight transient dynamics

We propose that *'epidemiology and evolution'* studies focus on demogenetic trajectories by taking demographic fluctuations of pathogen populations into account more systematically when studying plant resistance durability. To best characterise evolutionary trajectories, we have to consider the interplay between epidemiology and evolution, and treat them as having equal footing. Evolutionary forces shape epidemiological dynamics (for example when a new virulence leads to drastic epidemics on the resistant host). These population dynamics in turn impact population evolution (for example, because the size of the pathogen population influences the probability that a new mutation occurs).

Box 2: A brief history of population genetics

Population genetics is a research field with a long and productive history. Our goal here is not to draw a detailed picture of this area. Rather, we summarise the main points of interest to the reader from the viewpoint of this Opinion Note. Population genetics is a key element in evolutionary synthesis, because it formally explains the evolution of all heritable biological features for any defined transmission rule. The term 'population genetics' was coined nearly one century ago, most famously by Fisher (1930), Wright (1930) and Haldane (1924–1934). At its origin, the main contribution of this field was to show that a phenotype, even if distributed continuously, can be explained by the contribution of discrete entities called genes (Fisher, 1930). Nowadays, population genetics provides a valuable framework to fill the gap between genetic variation at an individual level (the genotypic state, *i.e.* a given combination of alleles, that determines a phenotype) and the change in this genetic variation at the population level. As such, it provides us with a sound approach for investigating molecular evolution. It also accounts for changes in genetic variation at different scales of time and space. Its applications range from point molecular variations to very complex genetic architectures as long as transmission rules from generation to generation are known (Ewens, 2004). More recently, population genetics have explicitly incorporated population dynamics in a so-called demogenetic approach, and hence is able to track feedback between epidemiology and evolution (see Section 4.1). Lastly, population genetics reasoning can be extended to multiple species to monitor their joint fate in time and space (Mueller 2019). For example, in studying host–plant interactions, population genetics allows analyses of the reciprocal selection pressures exerted by host and pathogen species and hence reveals their joint evolutionary trajectories (Brown & Tellier, 2011).

Population genetics constitutes a valuable approach as it considers the changes in both genotype states and numbers (Box 2), thus capturing this interplay between epidemiology and evolution. Such a modelling framework has been proposed by Day & Proulx (2004) and Day & Gandon (2007); notably, the latter study showed the existence of transient states that can be overlooked in studies focusing on final evolutionary equilibria only. These transient states are expected to emerge after a perturbation and can have strong epidemiological and evolutionary consequences. For example, the transient dynamics resulting from the onset of vaccination can lead to severe epidemics and the possible emergence of more virulent strains before epidemiological dynamics reach a new equilibrium (Gandon & Day, 2007).

The evolutionary epidemiology framework is worth applying to the management of plant resistances. Agricultural landscapes change rapidly, resulting in regular perturbations being applied to pathogen populations (Stukenbrock & McDonald, 2008; Zhan *et al.*, 2015). In the literature, we observe a change in paradigm from the short-term control of pathogen populations by means of epidemiology-driven regulation to the long-term control of the evolutionary trajectories of pathogen populations (Brown, 2015; Papaïx *et al.*, 2018). We thus need a proper definition of resistance durability that embraces both short- and long-term control of pathogen populations. Indeed, the initial definition according to Johnson (1979), "a resistance remaining effective in a cultivar for a long period of time during its widespread cultivation", is based on an arbitrary criterion. A more precise definition of resistance durability goes back to the study of van den Bosch & Gilligan (2003). This latter study was distinctive as it proposes several measures of resistance durability as well as a modelling framework coupling a classical epidemiological model to a population genetics approach. Hence, whereas some of their criteria for resistance durability are related to epidemic intensities, such as healthy area duration, other criteria, such as the time elapsing before a virulent allele to become fixed in the pathogen population, are typically derived from population genetics. Later, theoretical studies by Rimbaud *et al.* (2018a) and Papaïx *et al.* (2018) built on this distinction and proposed criteria related to epidemiological control and criteria related to evolutionary control. Importantly, they highlighted that strategies aiming to maximise epidemiological control are not necessarily those maximising evolutionary control (Rimbaud *et al.*, 2018a; Papaix *et al.*, 2018). Typically, landscapes planted with a high proportion of the resistant cultivar in low aggregation confer the best short-term epidemiological control, but are far from optimal for resistance durability (Papaix *et al.*, 2018).

Another interesting feature of the evolutionary epidemiology framework is that, with the same mathematical formalism, it can handle the dynamics of adaptation of pathogen populations towards both qualitative and quantitative resistances (Lo Iacono *et al.*, 2012; Fabre *et al.*, 2022). Furthermore, this formalism allows the dynamics of pathogen life history traits targeted by quantitative resistances to be derived (Lo Iacono *et al.*, 2012). Yet, this framework has seldom been applied in theoretical studies. One reason for this might be that it can lead to complex mathematical models, especially when dealing with quantitative resistances (Fabre *et al.*, 2022). Still, new models have to be developed, even considering the simpler genetic determinisms that govern qualitative interactions.

4.2 Build new contextual models that include detailed understanding of biological processes

We propose that '*epidemiology and evolution*' studies develop more realistic models of pathogen evolution. Pathogen genetics allows us to integrate the molecular mechanisms underlying pathogen adaptation through an explicit description of the genotypic states and the probabilities of their

changes. Losses of resistance efficacy can result from a variety of molecular mechanisms leading to alteration of the avirulence gene (Box 1). Knowledge of these key biological processes is of paramount importance to feed into epidemiological models and, ultimately, to forecast the risk of pathogen adaptation (Hessenauer *et al.*, 2021; Milgroom & Peever, 2003). Some mechanisms have already been taken into account in theoretical studies. For example, models consider both qualitative and quantitative resistances as general features, although quantitative resistances remain underrepresented (see Table 1 in Rimbaud *et al.*, 2021). However, our current knowledge of molecular mechanisms and pathogen biology ranges far beyond current modelling assumptions. We propose that modellers build contextual models that integrate our current understanding of biological processes. These models would be less generic but would lead to more accurate predictions on given classes of organisms. The following four examples illustrate the importance of considering such a wider array of biological processes underlying pathogen adaptation.

- (1) Life history traits affected by quantitative resistances drive resistance durability. Contrary to qualitative resistances that prevent infection, quantitative resistances can affect different steps of the infection cycle without completely preventing infection. Rimbaud *et al.* (2018b) predicted that resistance durability would differ depending on the trait impacted by quantitative resistance genes. In particular, this theoretical study showed that quantitative resistance targeting the pathogen latent period has the best potential to increase resistance durability. Similarly, evolution can lead in theory to a monomorphic or dimorphic pathogen population, depending on the life history trait impacted by quantitative resistance (sporulation curve versus infection efficiency) (Fabre *et al.*, 2022). This result highlights the importance of the choice of the quantitative resistance genes operated by plant breeders as a driver of pathogen diversification (Fabre *et al.,* 2022).

- (2) Molecular mechanisms rule pathogen adaptation. For example, the type and number of mutations is key to understanding pathogen adaptation. Population genetics allows us to set the probabilities of genotypic changes and to monitor pathogen evolution at the finest molecular scale. In the case of virus adaptation to qualitative resistances, the epidemiological control provided by qualitative resistance to a plant virus is impacted strongly by the number of mutations that the virus must accumulate to overcome resistance (Janzac *et al*., 2009; Fabre *et al.*, 2012; Djidjou-Demasse *et al.*, 2017). Here, typically, a population genetic concept, the mutation-selection equilibrium, allows molecular data (Janzac *et al*., 2009) and an evolutionary-epidemiological model (Fabre *et al.*, 2012; Djidjou-Demasse *et al.*, 2017) to be linked. Furthermore, few theoretical models consider epistatic interaction among avirulence gene loci, although molecular investigations have proved that it can occur (Plissonneau *et al.*, 2016). Finally, consider the example of priming, that is, the induction of plant defence by specific

infection conferring cross-protection to all pathogen genotypes. A theoretical study by Clin et al. (2021) found that considering priming changes predictions for the use of cultivar mixtures. With priming, a given proportion of resistant and susceptible host genotypes is optimal to reduce disease prevalence, whereas, without priming, maximal reduction is obtained with a monoculture of pure resistant plants.

- (3) The diverse, and sometimes complex, reproductive modes of pathogens are seldom considered. In particular, models often neglect the role of sexual reproduction (Rimbaud *et al.*, 2021). Yet, pathogens undergoing sexual reproduction generate, in theory, less gene diversity but greater evenness in genotypic distributions than asexual pathogens (Frank, 1993; Birky, 1996; Halkett *et al.*, 2005; Reichel et al., 2016; Stoeckel et al., 2021). Besides, mixed sexual and asexual reproductions could lead to higher risk of pathogen adaptation via an increased evolutionary potential: on the one hand sexual reproduction generates genotypic diversity and on the other hand asexual reproduction maintains and favors the most adapted combinations of alleles (McDonald & Linde, 2002; Bazin et al., 2014).

- (4) Few evolutionary epidemiology models take into account pathogen ploidy. Most studies assume haploid pathogens. Yet, diploid pathogens can follow different evolutionary trajectories. Following the gene-for-gene model, the recessivity of the virulence allele in diploids results in delayed and more stochastic invasion dynamics (see the modelling study by Saubin *et al.*, 2021). Moreover, pathogen life cycle variability can impact the relative importance of each evolutionary force on the resulting resistance durability. In particular host alternation in pathogen life cycle can reduce the probability of virulence emergence, but - once a virulence emerges - it can speed up resistance overcoming because host alternation increases gene flow (Saubin *et al.,* 2021).

More broadly, we argue that there are plenty of other molecular mechanisms and biological processes that need to be better considered in theoretical studies. This would lead to more contextualised and accurate studies of resistance durability.

4.3 Do not neglect genetic drift as a source of stochasticity in evolutionary trajectories

The diversity of pathogen life cycles translates into diversity in pathogen demography and population sizes. Small population sizes at any time in a pathogen life cycle increase genetic drift and, hence, stochasticity in evolutionary trajectories. We expect drift to impede pathogen adaptation, a weakness we can exploit to enhance resistance durability. The most recent theoretical models consider the stochasticity of pathogen evolutionary trajectories thanks to demogenetic models. However, few modelling works consider the effects of genetic drift (Rimbaud *et al.*, 2021). For some pathogen species, effective population sizes are shown to be large enough to confidently neglect genetic drift.

This is the case for many fungal species (e.g. *Botrytis cinerea*, Walker *et al.*, 2017; *Zymoseptoria tritici,* Zhan *et al.*, 2001). Other organisms have smaller population sizes and their evolution can be impacted more strongly by stochasticity: in particular, recent experimental studies highlighted the non-negligible effect of genetic drift on the dynamics of adaptation of diverse pathogens (e.g. in nematodes, Jan *et al.*, 2016; Montarry *et al.*, 2019).

Besides, some pathogen populations undergo large demographic fluctuations. Among the various pathogen life cycles, specific phases induce drastic drops in population size (*i.e.* bottlenecks). These include, for example, overwintering (Castel *et al.*, 2014; Djidjou-Demasse *et al.,* 2017; Fabre *et al.,* 2012; Fabre *et al.,* 2015; Hamelin *et al.,* 2011, Lo Iacono *et al.*, 2013), host alternation (Saubin et al., 2021) and plant-to-plant transmission (Moury et al., 2007). Quantitative resistances can also promote drift by reducing within-host population size (Rousseau *et al.,* 2017). Drift should therefore be considered more in modelling works (e.g. as in Rousseau *et al.*, 2019; Saubin *et al.*, 2021), as a direct consequence of demographic fluctuations, and in breeding processes as a leverage to slow pathogen adaptation (Quenouille *et al.*, 2014; Montarry *et al.*, 2019). In that respect, it could also be worth considering evolutionary processes in which genetic drift is particularly strong, such as reproductive modes that impact effective population sizes and bottleneck events. For example, during bottleneck events, evolutionary rescue can occur, when genetic adaptation prevents population extinction and leads to demographic recovery (Alexander *et al.*, 2014). This would shed new light on the effect of landscape management strategies on changes in pathogen population sizes in relation to their dynamics of adaptation.

However, we identify some limits to this proposed leverage. First, selection can occur at the parcel scale rather than within-host (Bergstrom, 1999). This is the case when there is recurrent migration events among hosts (horizontal transmission), which increases the effective size of the pathogen population. This challenges the possibility to take advantage of genetic drift by deploying quantitative resistances. Secondly, the increase in genetic drift would lead to more stochastic pathogen evolutionary trajectories, making it more difficult to forecast pathogen adaptation and to adapt landscape management accordingly.

4.4 Take into account the temporal dimension in empirical studies

Besides detailed investigations on molecular interactions, empirical studies should consider temporal samples to better elucidate rapid changes in molecular weaponries. Pathogens offer the unique opportunity to study contemporaneous evolution (McDonald, 2010) and to disentangle the interplay of the evolutionary forces at work. The recent advances in molecular studies promoted by omic

approaches have helped deciphering the diversity of effectors and the multiplicity of evolutionary mechanisms that can lead to the emergence of virulence (Fouché *et al.*, 2018; Rouxel & Balesdent, 2017; Stergiopoulos *et al.*, 2007; Stukenbrock & Bataillon, 2012). Detailed studies have examined the diversity of virulence determinants within pathogen populations (Daverdin *et al.*, 2012; Huang *et al.*, 2014) or evolutionary patterns in effector diversification (Stukenbrock & McDonald, 2009). Yet, a temporal dimension in pathogen population surveys is often lacking, or is too narrow to trace evolutionary dynamics in detail, as has been done using phylogenomic approaches in human diseases (Geoghegan & Holmes, 2018). Examining genetic variation on a snapshot sample can give access to only a restricted part of the evolutionary history and hides phenomena that remain unsuspected without the use of temporal samples (e.g. population replacement, as evidenced in Persoons *et al.*, 2017 and Louet *et al.*, 2022). We encourage long-term sampling efforts to build large collections of isolates over extended time periods (Gandon *et al.*, 2016).

4.5 Use the determinants of reciprocal interactions to forecast and stabilise evolutionary trajectories of pathogen populations

Finally, we propose better consideration of the reciprocity of interactions between hosts and pathogens within an evolutionary framework. In agroecosystems, host populations and their genetic diversity is strictly controlled by humans. As such, it is not correct to draw analogies between evolution in agroecosystems and the coevolution that occurs in nature. Yet, we argue that an understanding of the theoretical outcomes of host–pathogen coevolution can provide insights into the best ways to manage such host diversity for durable control of pathogen populations. Theoretical population genetics models show that coevolutionary dynamics can stabilise (or not) polymorphism, depending on how evolution of pathogen and host populations is coupled (Tellier & Brown, 2011). Strong and reciprocal interactions lead to unstable spirals, leading to evolutionary arms races through epidemic outbreaks. These occur when pathogen fitness depends on the abundance of a host plant that imposes directional selection. This is often typically observed in human-managed systems, with boom and bust cycles resulting from the strong and unidirectional selection pressures exerted by the intensive use of a single qualitative resistance (Zhan *et al.*, 2015; Persoons *et al*. 2017; Phan *et al*., 2020). A rupture of continuous selection pressure can enhance resistance durability, as observed for deployment of the *mlo* gene in barley against powdery mildew. A striking feature of this deployment is the long-term and lasting choice to refrain from introducing it in winter barley (Dreiseitl 2020). The obligate powdery mildew pathogen is deemed to infect winter varieties to overwinter. Therefore, despite the massive use of the *mlo* gene in spring varieties for more than four decades, each year the pathogen populations experience an interruption of selection pressure on winter varieties. More generally, ecological

processes that maintain the genetic diversity of host and pathogen populations are expected to regulate epidemic cycles and prevent outbreaks (Laine & Barrès, 2013). This can occur, for example, in natural systems with perennial hosts or seed banks (Tellier & Brown, 2009). Host and pathogen diversity is a key element in explaining the success of cultivar mixtures in disease control, in both the short and long term (Clin *et al.*, 2021), as recently observed by Perronne *et al.* (2021) in wheat. We should therefore promote and use the diversity of plant material provided by plant breeders. More generally, diversifying crops in agricultural landscape would be beneficial (Vialatte et al., 2021). A remaining question is how to go beyond the epidemiological approach in integrating coevolutionary dynamics, to make best use of this diversity.

The theory of coevolution indicates that polymorphism stability depends on the costs associated with host resistance(s) and pathogen virulence(s). Even if the fitness costs of virulence have been documented in all pathogen kingdoms (Leach *et al.* 2001), it is not a strict and generalised rule (see e.g. Maupetit *et al.* 2021, Fournet *et al.*, 2016). Quantitative resistances offer the opportunity to exploit putative trade-offs among pathogen fitness traits to better stabilise the system (Laine & Barrès, 2013). Early theoretical models of plant–pathogen coevolution focussed on gene-for-gene interactions. New models incorporating the genetic architecture of quantitative traits in both plants and pathogens should be developed.

Ultimately, a theoretical understanding of host–pathogen evolutionary trajectories should enable us to forecast pathogen evolution conditionally on the selection exerted by crops. Direct epidemiological surveys can allow us to know the current and past pathogen population dynamics and genetic characteristics, and to adequately adapt cultivated landscapes. Understanding the reciprocity between the human-mediated evolutionary trajectories of hosts and pathogens would make it possible to tailor plant breeding not only to current, but also anticipated, states of evolution of pathogen populations, provided plant resistance is evaluated early enough in the breeding process. This understanding and prediction of pathogen evolutionary trajectories would then lead to more durable management strategies of host resistance, including adaptive adjustments of the varietal landscape. In addition to using all available genetic diversity developed by plant breeders to break uni-directional selection pressures, we advise strengthening plant breeding efforts by including pathologists, molecular biologists, and modellers as early as possible when developing breeding schemes.

Concluding remarks

Population genetics is a central framework that can improve our understanding of the durability of plant resistance, by knitting together '*plant breeding*', *'molecular interactions',* and '*epidemiology and evolution*' analyses. However, this research field is still underrepresented, both empirically and theoretically, and often used only as a simplified tool. We believe that enhanced dialogue across the three identified subfields will pave the way to a better knowledge of the evolutionary trajectories of pathogen populations and help improve the sustainable management of genetic resistances. Developing interdisciplinary work is challenging. It may be costly and risky in the short term, compared with traditional studies deeply rooted in each subfield. Incentives must, however, be strong enough to bypass these initial difficulties and ensure that teams with complementary expertise work together in a long-term effort. Ultimately, we advocate forming interdisciplinary teams by bringing together researchers from the three subfields and truly sharing the same goal—namely the development and deployment of durable resistances. Our prediction is that this will lead to significant progress in the sustainable management of plant diseases.

Acknowledgements

We thank Helen Rothnie for the correction of the English language. We also thank three anonymous reviewers and the associate editor for detailed comments and suggestions which greatly helped us to improve this Opinion Note. This work was supported by grants from the French National Research Agency (ANR-18-CE32-0001, Clonix2D project; ANR-11-LABX-0002-01, Cluster of Excellence ARBRE) and the Metaprogram SuMCrop of the National Research Institute for Agriculture, Food and the Environment (INRAE , Opiniâtres project). Clémentine Louet was supported by a PhD fellowship from the Région Lorraine and the French National Research Agency (ANR-18-CE32-0001, Clonix2D project). Méline Saubin was supported by a PhD fellowship from INRAE and the French National Research Agency (ANR-18-CE32-0001, Clonix2D project).

Data accessibility

All the files extracted from the ISI Web of Science portal (https://www.webofscience.com) and the list of keywords used for the bibliographic analysis are available on Zenodo repository (DOI: 10.5281/zenodo.6651957).

Author contributions

This Opinion Note originates from a collaborative project called "ENDURANCE" submitted to the French National Research Agency. M.S., C.L., B.P. and F.Hal. defined the search terms, extracted and cured the bibliographic corpus. M.S. performed network bibliographic analyses. M.S., C.L., B.P. and F.Hal. drafted the first version of this note. All authors contributed to the writing and revision and approved the manuscript.

References

Alexander, H. K., Martin, G., Martin, O. Y., Bonhoeffer, S. (2014). Evolutionary rescue: linking theory for conservation and medicine. *Evolutionary Applications*, *7*(10), 1161–1179.

Bazin, É., Mathé-Hubert, H., Facon, B., Carlier, J., Ravigné, V. (2014). The effect of mating system on invasiveness: Some genetic load may be advantageous when invading new environments. *Biological Invasions*, *16*(4), 875–886.

Bergstrom, C. T., McElhany, P., Real, L. A. (1999). Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proceedings of the National Academy of Sciences*, *96*(9), 5095–5100.

Birky, C. W. (1996). Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics*, 144, 427–437.

Brown, J. K. M. (2015). Durable resistance of crops to disease: a Darwinian perspective. *Annual Review of Phytopathology*, *53*, 513–539.

Brown, J. K. M., Tellier, A. (2011). Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual Review of Phytopathology*, *49*, 345–367.

Brun, H., Chèvre, A.-M., Fitt, B. D. L., Powers, S., Besnard, A.-L., Ermel, M., Huteau, V., Marquer, B., Eber, F., Renard, M., Andrivon, D. (2010). Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus. New Phytologist*, *185*(1), 285–299.

Castel, M., Mailleret, L., Andrivon, D., Ravigné, V., Hamelin, F. M. (2014). Allee effects and the evolution of polymorphism in cyclic parthenogens. *American Naturalist, 183*, E75–E88.

Clin, P., Grognard, F., Mailleret, L., Val, F., Andrivon, D., Hamelin, F. M. (2021). Taking advantage of pathogen diversity and immune priming to minimize disease prevalence in host mixtures: a model. *Phytopathology*, *111*(7), 1219–1227.

Cowger, C., Brown, J. K. M. (2019). Durability of quantitative resistance in crops: greater than we know? *Annual Review of Phytopathology*, *57*, 253–277.

Cowger, C., Mundt, C. C. (2002). Effects of wheat cultivar mixtures on epidemic progression of Septoria tritici blotch and pathogenicity of *Mycosphaerella graminicola*. *Phytopathology, 92*, 617–23.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., Varshney, R. K. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science. 22*, 961–975.

Dangl, J. L., Horvath, D. M., Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science, 341*(6147), 746–751.

Daverdin, G., Rouxel, T., Gout, L., Aubertot, J.-N., Fudal, I., Meyer, M., Parlange, F., Carpezat, J., Balesdent, M.-H. (2012). Genome structure and reproductive behaviour influence the evolutionary potential of a fungal phytopathogen. *PLOS Pathogens*, *8*(11), e1003020.

Day, T., & Gandon, S. (2007). Applying population-genetic models in theoretical evolutionary epidemiology. *Ecology Letters*, *10*(10), 876–888.

Day, T., Proulx, S. R. R. (2004). A general theory for the evolutionary dynamics of virulence. *The American Naturalist*, *163*(4), E40–E63.

Djidjou-Demasse, R., Moury, B., Fabre, F. (2017). Mosaics often outperform pyramids: insights from a model comparing strategies for the deployment of plant resistance genes against viruses in agricultural landscapes. *New Phytologist*, *216*(1), 239–253.

Dreiseitl, A. (2020). A novel way to identify specific powdery mildew resistance genes in hybrid barley cultivars. *Sci Rep 10*, 18930.

Ewens, W. J. (2004). Mathematical population genetics 1: theoretical introduction. *Springer Science & Business Media*, 27.

Fabre, F., Rousseau, E., Mailleret, L., Moury, B. (2012). Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytologist*, *193*(4), 1064–1075.

Fabre, F., Rousseau, E., Mailleret, L., Moury, B. (2015). Epidemiological and evolutionary management of plant resistance: Optimizing the deployment of cultivar mixtures in time and space in agricultural landscapes. *Evolutionary Applications, 8*, 919-932.

Fabre, F., Burie, J.-B., Ducrot, A., Lion, S., Richard, Q., Djidjou-Demasse, R. (2022). An epi-evolutionary model for predicting the adaptation of spore-producing pathogens to quantitative resistance in heterogeneous environments. *Evolutionary Applications*, *15*, 95– 110.

Fisher, R. A. (1930). The genetical theory of natural selection. *Oxford: Clarendon Press.*

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, *9*, 275–296.

Fouché, S., Plissonneau, C., Croll, D. (2018). The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Current Opinion in Microbiology*, *46*, 34–42.

Fournet, S., Eoche-Bosy, D., Renault, L., Hamelin, F. M., Montarry, J. (2016). Adaptation to resistant hosts increases fitness on susceptible hosts in the plant parasitic nematode *Globodera pallida*. *Ecology and evolution*, *6*(8), 2559–2568.

Fournet, S., Kerlan, M. C., Renault, L., Dantec, J. P., Rouaux, C., Montarry, J. (2013). Selection of nematodes by resistant plants has implications for local adaptation and cross-virulence. *Plant Pathology*, *62*(1), 184–193.

Frank, S.A. (1993). Coevolutionary genetics of plants and pathogens. *Evol Ecol,* 7, 45–75.

Gandon S., Day, T. (2007). The evolutionary epidemiology of vaccination. *J R Soc Interface*, *4*(16), 803–17.

Gandon, S., Day, T., Metcalf, C. J. E., Grenfell, B. T. (2016). Forecasting epidemiological and evolutionary dynamics of infectious diseases. *Trends in Ecology and Evolution, 31*(10), 776-788.

Gandon, S., Michalakis, Y. (2000). Evolution of parasite virulence against qualitative or quantitative host resistance. *Proc. R. Soc. Lond. B*, *267*, 985–990.

García-Arenal, F., Fraile, A., Malpica, J. M. (2003). Variation and evolution of plant virus populations. *International Microbiology*, *6*(4), 225–232.

Geoghegan, J. L., & Holmes, E. C. (2018). The phylogenomics of evolving virus virulence. *Nature Reviews Genetics*, *19*(12), 756-769.

Gómez, P., Rodríguez-Hernández, A. M., Moury, B., Aranda, M. A. (2009). Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. *European Journal of Plant Pathology*, *125*(1), 1–22.

Haldane, J. B. S. (1924-1934). A mathematical theory of natural and artificial selection I-X. *Bulletin of Mathematical Biology*, *52*, 209–240.

Halkett, F., Simon, J., & Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, *20*(4), 194–201.

Hamelin, F. M., Castel, M., Poggi, S., Andrivon, D., Mailleret, L. (2011). Seasonality and the evolutionary divergence of plant parasites. *Ecology, 92*, 2159–2166.

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *The ISME journal, 11*(5), 1189–1204.

Hessenauer, P., Feau, N., Gill, U., Schwessinger, B., Brar, G. S., Hamelin, R. C. (2021). Evolution and adaptation of forest and crop pathogens in the Anthropocene. *Phytopathology*, *111*(1), 49–67.

Huang, Y. J., Li, Z. Q., Evans, N., Rouxel, T., Fitt, B. D. L., Balesdent, M. H. (2006). Fitness cost associated with loss of the AvrLm4 avirulence function in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). *European Journal of Plant Pathology*, *114*(1), 77–89.

Huang, J., Si, W., Deng, Q., Li, P., Yang, S. (2014). Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. BMC Genetics, *15*(1), 45.

Jan, P.-L., Gracianne, C., Fournet, S., Olivier, E., Arnaud, J.-F., Porte, C., Bardou-Valette, S., Denis, M.-C., Petit, E. J. (2016). Temporal sampling helps unravel the genetic structure of naturally occurring populations of a phytoparasitic nematode. 1. Insights from the estimation of effective population sizes. *Evolutionary Applications*, *9*(3), 489–501.

Janzac, B., Fabre, F., Palloix, A., Moury, B. (2009). Constraints on evolution of virus avirulence factors predict the durability of corresponding plant resistances. *Molecular Plant Pathology*, *10*(5), 599–610.

Johnson, R. (1984). A critical analysis of durable resistance. *Annual Review of Phytopathology*, *22*, 309–330.

Jones, J. D. G., Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329.

Jørgensen, I. H. (1992). Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. *Euphytica*, *63*(1), 141-152.

Kusch, S., Panstruga, R. (2017). mlo-based resistance: an apparently universal "weapon" to defeat powdery mildew disease. *Mol Plant Microbe Interact*, *30*(3), 179-189.

Laine, A.-L., Barrès, B. (2013). Epidemiological and evolutionary consequences of life-history trade-offs in pathogens. *Plant Pathol, 62*, 96–105.

Laloi, G., Vergne, E., Durel, C. E., Le Cam, B., Caffier, V. (2017). Efficiency of pyramiding of three quantitative resistance loci to apple scab. *Plant Pathology*, *66*(3), 412-422.

Leach, J. E., Vera Cruz, C. M., Bai, J., Leung, H. (2001). Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annual Review of Phytopathology*, *39*(1), 187–224.

Lo Iacono, G., van den Bosch, F., Gilligan, C. A. (2013). Durable resistance to crop pathogens: an epidemiological framework to predict risk under uncertainty. *PLOS Computational Biology, 9*(1): e1002870.

Lo Iacono, G., van den Bosch, F., Paveley, N. (2012). The evolution of plant pathogens in response to host resistance: Factors affecting the gain from deployment of qualitative and quantitative resistance. *Journal of Theoretical Biology*, *304*, 152–163.

Louet, C., Saubin, M., Andrieux, A., Persoons, A., Gorse, M., Pétrowski, J., Fabre, B., De Mita, S., Duplessis, S., Frey, P., Halkett, F. (2022). A point mutation and large deletion at the candidate avirulence locus *AvrMlp7* in the poplar rust fungus correlate with poplar RMlp7 resistance breakdown. *Molecular Ecology*, *00*, 1– 12.

Lyngkjær, M. F., Newton, A. C., Atzema, J. L., Baker, S. J. (2000). The Barley *mlo*-gene: an important powdery mildew resistance source. *Agronomy*, *20*(7), 745-756.

Maupetit, A., Fabre, B., Pétrowski, J., Andrieux, A., De Mita, S., Frey, P., Halkett, F., Hayden, K. (2021). Evolution of morphological but not aggressiveness-related traits following a major resistance breakdown in the poplar rust fungus, *Melampsora larici-populina*. *Evolutionary Applications*, *14,* 513–523.

McDonald, B. (2010). How can we achieve durable disease resistance in agricultural ecosystems? *New Phytologist*, *185*(1), 3–5.

McDonald, B. A., Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, *40*(1), 349–379.

Milgroom, M. G., Peever, T. L. (2003). Population biology of plant pathogens: the synthesis of plant disease epidemiology and population genetics. *Plant Disease*, *87*(6), 608–617.

Montarry, J., Bardou-Valette, S., Mabon, R., Jan, P.-L., Fournet, S., Grenier, E., Petit, E. J. (2019). Exploring the causes of small effective population sizes in cyst nematodes using artificial *Globodera pallida* populations. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1894), 20182359.

Moury, B., Fabre, F., Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. *Proc. Natl. Acad. Sci. 104*:17891–17896.

Mueller, L. (2019). Conceptual breakthroughs in evolutionary ecology. *Academic press, Elsevier*.

Nilusmas, S., Mercat, M., Perrot, T., Djian-Caporalino, C., Castagnone-Sereno, P., Touzeau, S., Calcagno, V., Mailleret, L. (2020). Multiseasonal modelling of plant-nematode interactions reveals efficient plant resistance deployment strategies. *Evolutionary Applications*, *13*(9), 2206–2221.

Nováková, M., Šašek, V., Trdá, L., Krutinová, H., Mongin, T., Valentová, O., Balesdent, M.-H., Rouxel, T., Burketová, L. (2016). *Leptosphaeria maculans* effector AvrLm4-7 affects salicylic acid (SA) and ethylene (ET) signalling and hydrogen peroxide (H2O2) accumulation in *Brassica napus*. *Molecular Plant Pathology*, *17*(6), 818–831.

Palloix, A., Ayme, V., Moury, B. (2009). Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. *New Phytologist*, *183*(1), 190–199.

Papaïx, J., Rimbaud, L., Burdon, J. J., Zhan, J., Thrall, P. H. (2018). Differential impact of landscape-scale strategies for crop cultivar deployment on disease dynamics, resistance durability and long-term evolutionary control. *Evolutionary Applications*, *11*(5), 705–717.

Pariaud, B., Ravigné, V., Halkett, F., Goyeau, H., Carlier, J. and Lannou, C. (2009). Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathology*, *58*, 409–424.

Perronne, R., Dubs, F., de Vallavieille-Pope, C., Leconte, M., du Cheyron, P., Cadot, V., Vidal, T., Enjalbert, J. (2021). Spatiotemporal changes in varietal resistance to wheat yellow rust in France reveal an increase in field resistance level during the period 1985–2018. *Phytopathology*, *111*(9), 1602–1612.

Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, *26*(7), 1902–1918.

Petit-Houdenot, Y., Fudal, I. (2017). Complex interactions between fungal avirulence genes and their corresponding plant resistance genes and consequences for disease resistance management. *Frontiers in Plant Science*, *8*(1072).

Phan, H. T., Jones, D. A., Rybak, K., Dodhia, K. N., Lopez-Ruiz, F. J., Valade, R., Gout, L., Lebrun, M.-H., Brunner, P. C., Olivier, R. P., Tan, K.-C. (2020). Low amplitude boom-and-bust cycles define the Septoria nodorum blotch interaction. *Frontiers in Plant Science*, *10*, 1785.

Piffanelli, P., Ramsay, L., Waugh, R., Benabdelmouna, A., D'Hont, A., Hollricher, K., Jørgensen, J. H., Schulze-Lefert, P., Panstruga, R. (2004). A barley cultivation-associated polymorphism conveys resistance to powdery mildew. *Nature*, *430*, 887–91.

Plissonneau, C., Daverdin, G., Ollivier, B., Blaise, F., Degrave, A., Fudal, I., Rouxel, T., Balesdent, M.-H. (2016). A game of hide and seek between avirulence genes *AvrLm4-7* and *AvrLm3* in *Leptosphaeria maculans*. *New Phytologist*, *209*(4), 1613–1624.

Quenouille, J., Paulhiac, E., Moury, B., Palloix, A. (2014). Quantitative trait loci from the host genetic background modulate the durability of a resistance gene: a rational basis for sustainable resistance breeding in plants. *Heredity, 112*, 579 – 5 87.

Reichel, K., Masson, J.-P., Malrieu, F., Arnaud-Haond, S., Stoeckel, S. (2016). Rare sex or out of reach equilibrium? The dynamics of FIS in partially clonal organisms. *BMC Genetics*, 17(76).

REX Consortium (2010). The skill and style to model the evolution of resistance to pesticides and drugs: Modelling resistance evolution. *Evolutionary Applications*, *3*(4), 375–390.

REX Consortium (2016). Combining selective pressures to enhance the durability of disease resistance genes. *Frontiers in Plant Science*, *7,* 1916.

Ribaut, J. M., Hoisington, D. (1998). Marker-assisted selection: new tools and strategies. *Trends in Plant Science*, *3*(6), 236–239.

Rimbaud, L., Fabre, F., Papaïx, J., Moury, B., Lannou, C., Barrett, L. G., Thrall, P. H. (2021). Models of plant resistance deployment. *Annual Review of Phytopathology*, *59*(1).

Rimbaud, L., Papaïx, J., Barrett, L. G., Burdon, J. J., Thrall, P. H. (2018a). Mosaics, mixtures, rotations or pyramiding: What is the optimal strategy to deploy major gene resistance? *Evolutionary Applications*, *11*(10), 1791–1810.

Rimbaud, L., Papaïx, J., Rey, J.-F., Barrett, L. G., Thrall, P. H. (2018b). Assessing the durability and efficiency of landscape-based strategies to deploy plant resistance to pathogens. *PLOS Computational Biology*, *14*(4), e1006067.

Rousseau, E., Bonneault, M., Fabre, F., Moury, B., Mailleret, L., Grognard, F. (2019). Virus epidemics, plant-controlled population bottlenecks and the durability of plant resistance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1775), 20180263.

Rousseau, E., Moury, B., Mailleret, L., Senoussi, R., Palloix, A., Simon, V., Valière, S., Grognard, F., Fabre, F. (2017). Estimating virus effective population size and selection without neutral markers. *PLOS Pathogens*, *13*(11), e1006702.

Rouxel, T., Balesdent, M. H. (2017). Life, death and rebirth of avirulence effectors in a fungal pathogen of Brassica crops, *Leptosphaeria maculans*. *New Phytologist*, *214*(2), 526–532.

Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F. E., Soyer, J. L., Tellier, A., Croll, D. (2018). The genome biology of effector gene evolution in filamentous plant pathogens. *Annual Review of Phytopathology*, *56*, 21–40.

Saubin, M., Mita, S. De, Zhu, X., Sudret, B., Halkett, F. (2021). Impact of ploidy and pathogen life cycle on resistance durability. *Peer Community Journal, 1*: e8.

Stergiopoulos, I., De Kock, M. J. D., Lindhout, P., De Wit, P. J. G. M. (2007). Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. *Molecular Plant-Microbe Interactions*, *20*(10), 1271–1283.

Stoeckel, S., Arnaud-Haond, S., Krueger-Hadfield, S. A. (2021). The combined effect of haplodiplonty and partial clonality on genotypic and genetic diversity in a finite mutating population. *Journal of Heredity*, 112(1) 78–91.

Stukenbrock, E. H., Bataillon, T. (2012). A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. *PLoS Pathogens*, *8*(9), e1002893.

Stukenbrock, E. H., McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, *46*, 75-100.

Stukenbrock, E. H., McDonald, B. A. (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions*, *22*(4), 371–380.

Tellier, A., Brown, J. K. (2009). The influence of perenniality and seed banks on polymorphism in plant-parasite interactions. *American Naturalist, 174(6*), 769-79.

Tellier, A., Brown, J. K. (2011). Spatial heterogeneity, frequency-dependent selection and polymorphism in host-parasite interactions. *BMC Evol Biol **11,*** 319.

Vialatte, A., Tibi, A., Alignier, A., Angeon, V., Bedoussac, L., Bohan, D.A., Bougherara, D., Carpentier, A., Castagneyrol, B., Cordeau, S., Courtois, P., Deguine, J-P., Enjalbert, J., Fabre, F., Féménia, F., Fréville, H., Goulet, F., Grateau, R., Grimonprez, B., Gross, N., Hannachi, M., Jeanneret, P., Kuhfuss, L., Labarthe, P., Launay, M., Lefebvre, M., Lelièvre, V., Lemarié, S., Martel, G., Masson, G., Navarrete, M., Plantegenest, M., Ravigné, V., Rusch, A., Suffert, F., Tapsoba, A., Therond, O., Thoyer, S., Martinet, V. (2021). Chapter Four - Promoting crop pest control by plant diversification in agricultural landscapes: A conceptual framework for analysing feedback loops between agro-ecological and socio-economic effects. *Advances in Ecological Research, 65*, 133-165.

Van den Bosch, F., Gilligan, C. A. (2003). Measures of durability of resistance. *Phytopathology*, *93*(5), 616–625.

Wakeley, J. (2005). The limits of theoretical population genetics. *Genetics*, *169*(1), 1–7.

Walker, A.-S., Ravigne, V., Rieux, A., Ali, S., Carpentier, F., Fournier, E. (2017). Fungal adaptation to contemporary fungicide applications: the case of *Botrytis cinerea* populations from Champagne vineyards (France). *Molecular Ecology*, *26*(7), 1919–1935.

Wright, S. (1930). The genetical theory of natural selection: A review. *Journal of Heredity*, *21*(8), 349–356.

Zhan, J., Mundt, C. C., McDonald, B. A. (2001). Using Restriction Fragment Length Polymorphisms to Assess Temporal Variation and Estimate the Number of Ascospores that Initiate Epidemics in Field Populations of *Mycosphaerella graminicola*. *Phytopathology*, *91*(10), 1011–1017.

Zhan, J., Thrall, P. H., Papaïx, J., Xie, L., Burdon, J. J. (2015). Playing on a pathogen's weakness: using evolution to guide sustainable plant disease control strategies. *Annual Review of Phytopathology*, *53*(1), 19–43.

Annex to Improving sustainable crop protection using population genetics concepts: bibliographic analysis of the corpus of research articles on plant resistance durability

1 Introduction

This supplementary material details the bibliographic analysis performed in support of the opinion note entitled "Improving sustainable crop protection using population genetics concepts". We first indicate how we extracted and curated the corpus of articles referring to the durability of plant resistances. Then, we provide the reader with a step-by-step analysis of this corpus regarding the geographic distribution of research articles, the journals targeted, how the bibliographic corpus is structured, and how this structure evolved through time.

In this bibliographic analysis, we paid particular attention to the way population genetics is used in articles referring to plant resistance durability. We were especially interested in investigating the position of the 'population genetics' keyword in our corpus, as we hypothesized that this research field would provide a link between plant breeding, molecular and epidemiological articles of plant resistance durability.

2 Corpus selection

A representative corpus of research articles on plant resistance durability was extracted from the ISI Web of Science portal (https://www.webofscience.com) with the following request: 'Plant AND Resistan* AND Durab*'. A first corpus of 2522 articles was built considering all publication years for this extraction. This collection was then refined by categories to remove articles outwith the scope of our search (e.g. related to durable resistant materials for constructions). We also kept only articles cited at least once. The final corpus was

composed of 1783 articles from 1979 to 2021 (Fig. 1). This collection was exported with all metadata of the articles (e.g. authors, journal, country, times cited, keywords) referenced in the ISI Web of science. It is important to be aware that, despite the fact that all articles in the corpus studied do mention the combination of the three search terms, these articles did not necessarily state as their objective that they would exclusively tackle plant resistance durability. Moreover, articles reporting examples of durable plant resistance may be underrepresented owing to publication bias favouring the documentation of resistance overcoming (Cowger and Brown, 2019).

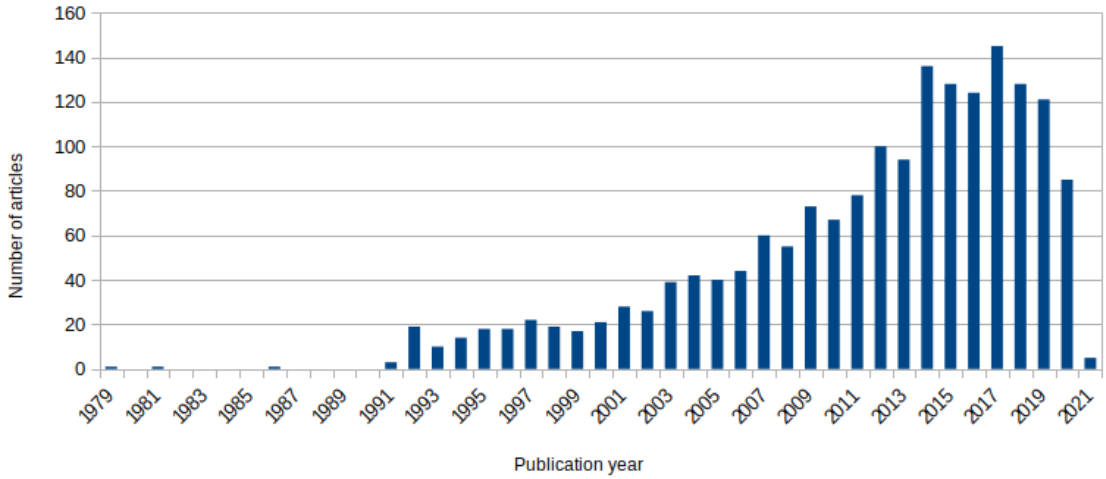


Fig. 1: Distribution of the number of articles in the corpus by year of publication.

All subsequent bibliographic analyses were performed with CorTexT Manager platform (https://www.cortext.net/) using the metadata of the articles. Structures and links among articles were displayed as network graphs. Nodes represent the metadata field(s) analysed (e.g. publication, keyword, country); one or two fields can be represented on the same network. Links between nodes are represented by lines, which indicate mutual occurrence of the field tagged on the node. The closer the nodes, the tighter the association. Nodes with high density of links among them are distinguished by colored circles and a main concept is associated with each cluster. All clustering analyses are based on the Louvain community detection algorithm, and node sizes are proportional to the number of co-occurrences they have with other nodes. We set all network mapping parameters to default values except for the number of nodes represented (subsequently specified).

3 Corpus analyses

3.1 Well-defined geographic research poles publish most of the corpus

To understand the geographical origins of articles in the corpus, a geographical analysis was based on the 'addresses' field to identify the number of articles originating from each region of the world. An article originated from a region when at least one author's address was located in that region. Articles could therefore originate from several regions depending on the authors' addresses. The analysis was performed with the default settings of the CorTexT Manager platform for geospatial exploration. Six countries contribute most to the corpus: United States, France, China, Australia, India, and United Kingdom (Fig. 2) with 22.0%, 11.8%, 10.8%, 6.1%, 5.8% and 4.3% of the corpus addresses, respectively.

To provide an outlook of the geographical co-authorship links between articles in the corpus, two co-authorship clustering analyses were performed on countries and research institutions, based on the 'addresses' or 'research institutions' field. Networks were obtained by representing 50 nodes for the countries and 70 nodes for the research institutions. Links and distances between nodes (countries or institutions) depend on the number of articles with co-authors originating from these two different nodes (countries or institutions). The co-citation network roughly matches the world map, with tighter links among neighbouring countries (Fig. 3). This is especially true for European countries. At a finer scale, this observation of co-citations linked to geographical proximity is strengthened by the frequent co-citations between national institutions (Fig. 4). In particular, French institutions form a distinct co-citation cluster in Fig. 4. This finer scale highlights research partnerships between specific institutions (e.g. CSIRO in Australia and Zurich university in Switzerland), which could explain further the clustering observed between countries. Altogether, these results reveal well-defined geographic researches and related institutions, with international collaborations strongly but not only influenced by geographical proximity.

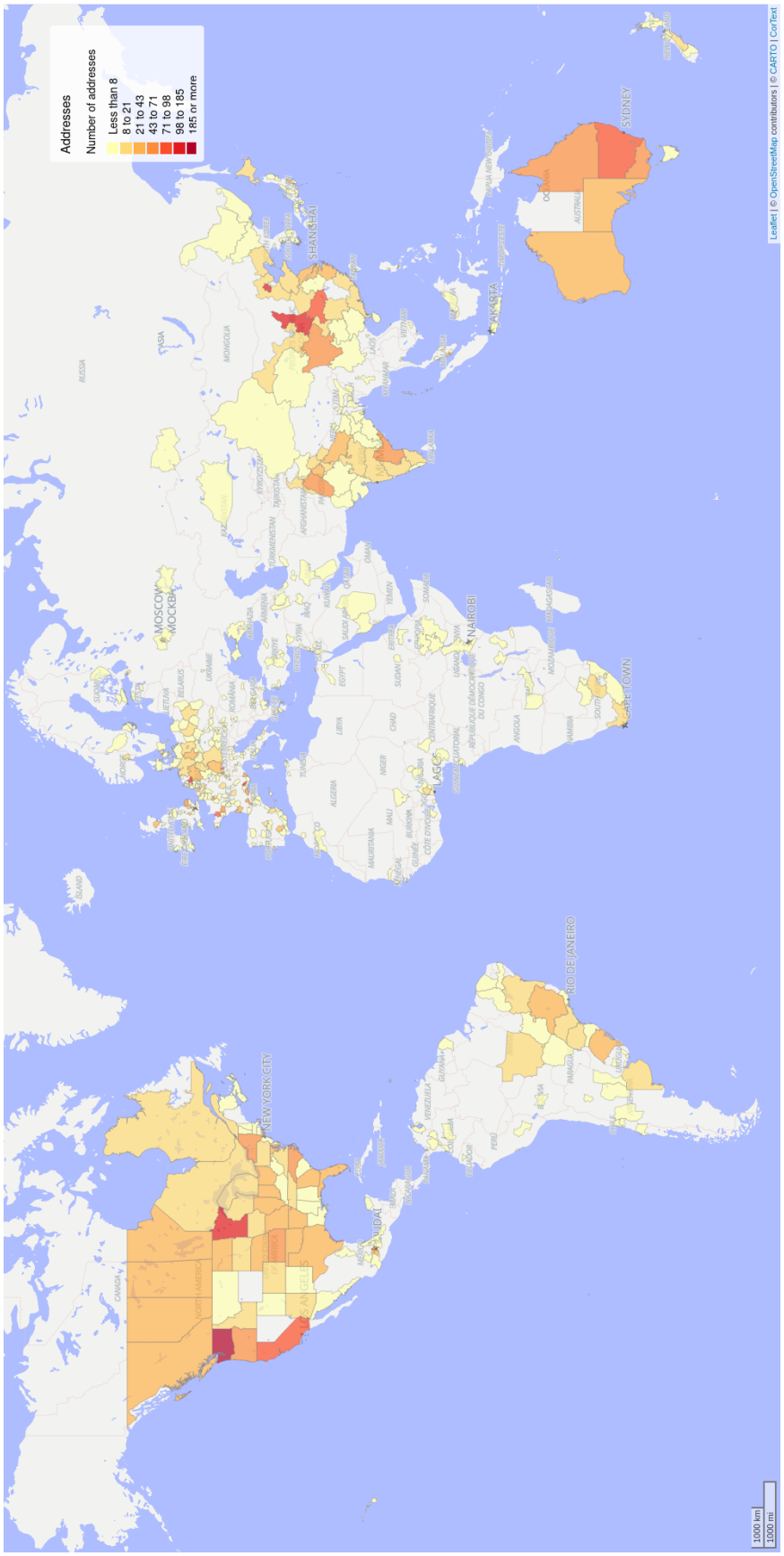


Fig. 2: Map representing the number of articles of the corpus originating from each region of the world. Regions are attributed to articles depending on the 'addresses' field referenced in the data base. This representation was drawn with the Geospatial Exploration tool on CorTexT Manager, with default parameters.

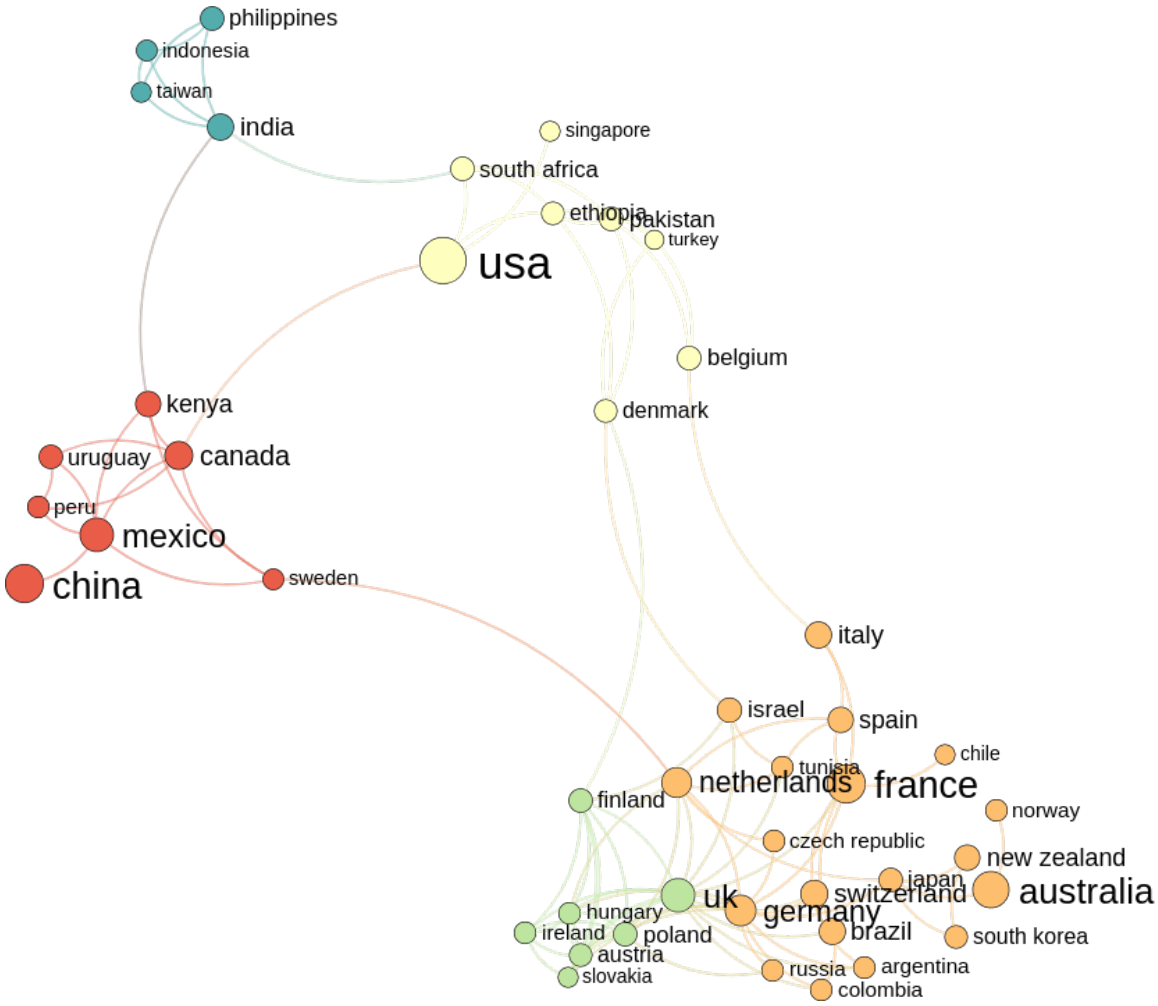


Fig. 3: Co-authorship network based on articles' countries. Countries are attributed to articles depending on the 'addresses' field referenced in the data base. This representation was drawn with the Network Mapping tool on CorTexT Manager, with 50 nodes and all other parameters set to default parameters.

3.2 Two main research communities emerge from the co-cited journal clustering

To highlight and understand links between fields of publications, a clustering analysis was performed on co-cited journals. This network represents 100 nodes and was based on the 'Cited Journal' field (Fig. 5). This clustering represents co-cited links between journals, which tend to be regrouped into two main research communities:

1. A molecular and mechanistic community (Plant Cell, Plant Physiology, Molecular Plant Microbe Interactions, Molecular Plant Pathology, etc.).

2. A phytopathology and plant breeding community (Annual Review of Phytopathology,

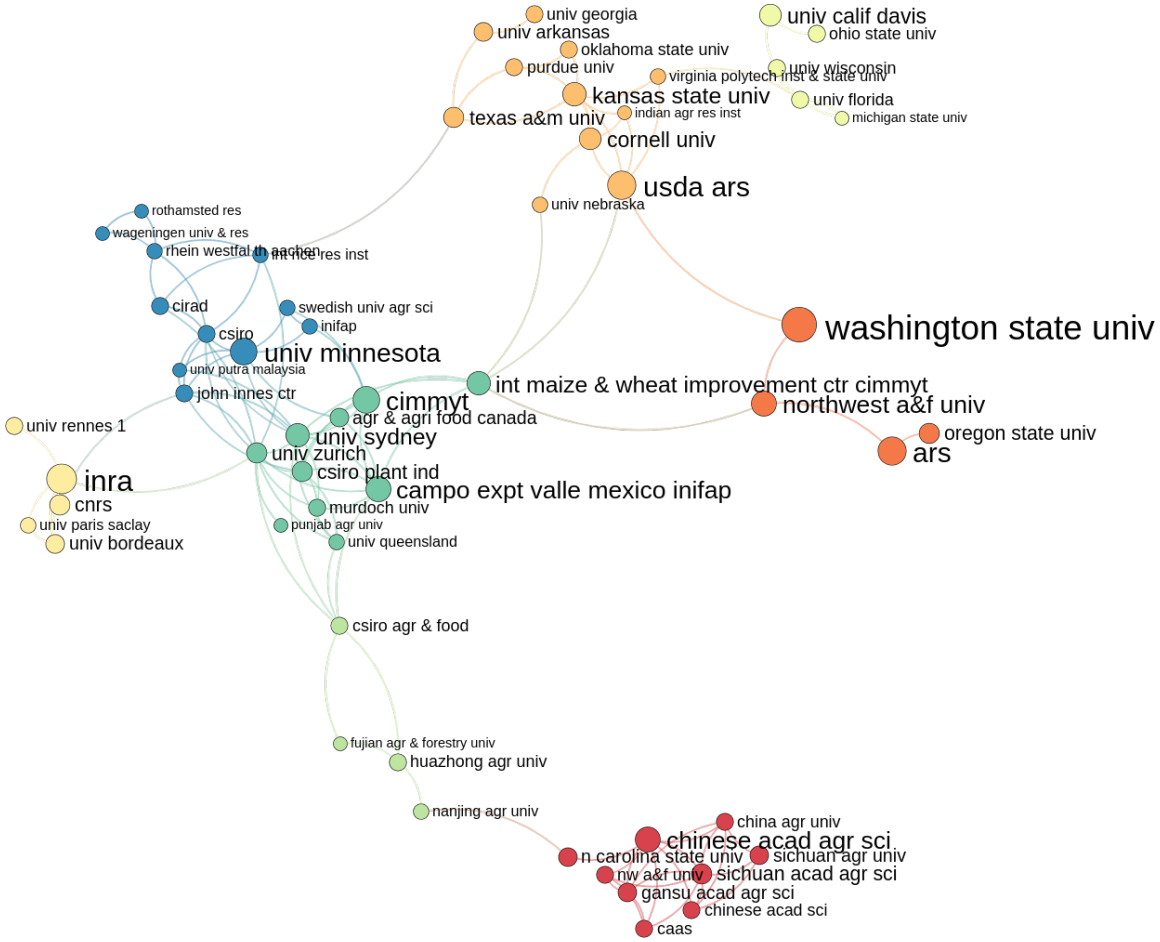


Fig. 4: Co-authorship network based on articles' institutions. Institutions are attributed to articles depending on the 'Research Institutions' field referenced in the data base. This representation was drawn with the Network Mapping tool on CorTexT Manager, with 70 nodes and all other parameters set to default parameters.

Phytopathology, Theoretical and applied genetics, etc.).

These two main clusters both associate plants and pathogens but not with the same research objectives. Apart from these two clusters, there are two small clusters dedicated to two specific biological models: viruses and insects, respectively.

3.3 The community is mainly structured by the model organisms studied

To identify the main research structures within the corpus, we base the following analyses on keywords illustrative of each article. A first extraction of keywords was performed, using CorTexT Manager platform, on article titles, abstracts, and WOS keywords. The later refer to automatic keyword referencing by the ISI Web of Science portal. The detection was based on nouns in the English language at the sentence level and ranked with chi2 scores. A network was performed based on the co-occurrence of 100 main structur-

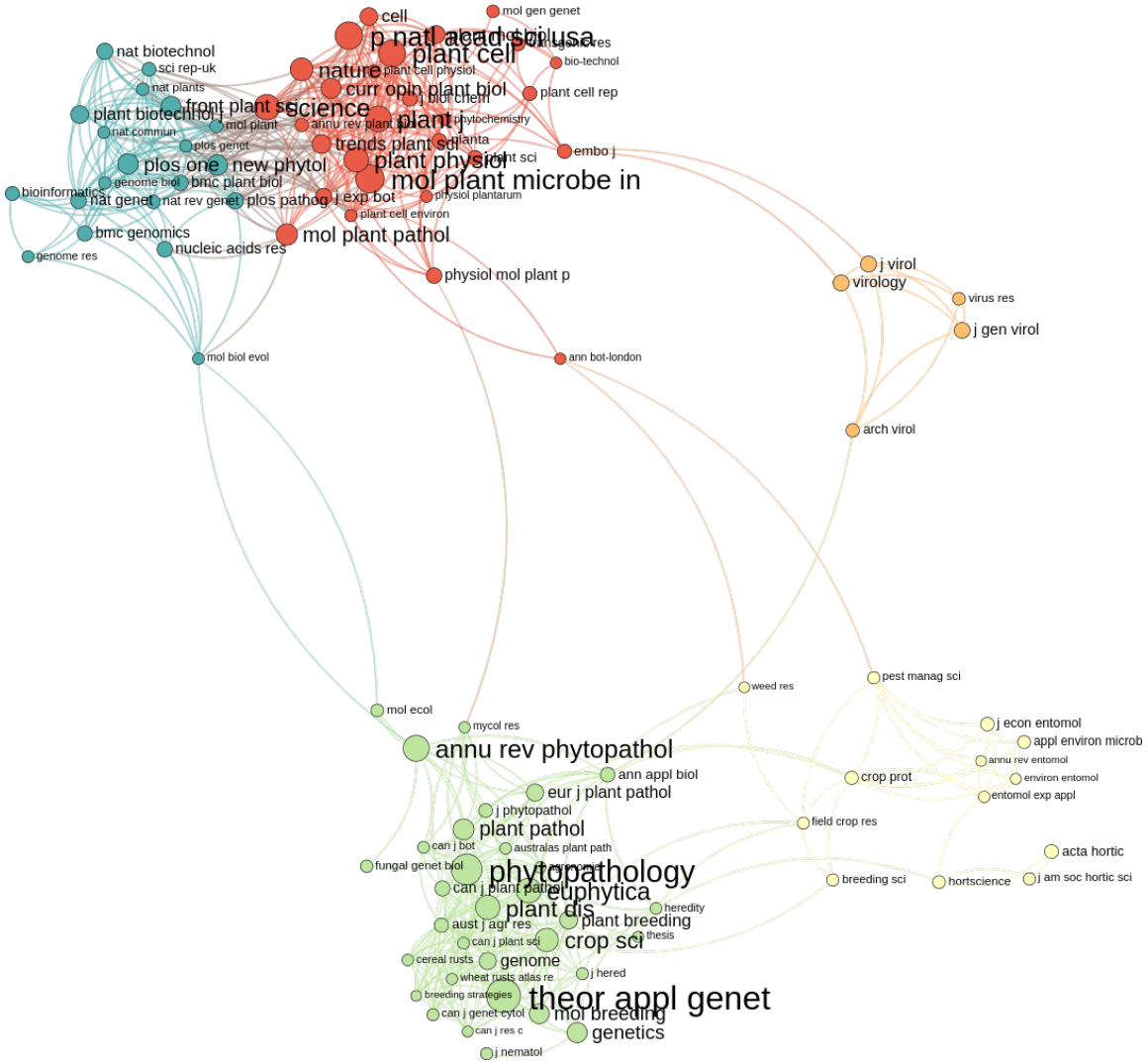


Fig. 5: Co-citation network based on the 'Cited Journal' field. This representation was drawn with the Network Mapping tool on CorTexT Manager, with 100 nodes and all other parameters set to default parameters.

ing keywords and the co-cited references, with 500 nodes represented. Thereby, for this clustering analyses, clusters regroup articles with similar keywords and cited references (Fig. 6). This clustering highlights a strong structuring of the corpus articles according to model organisms: of the ten clusters emerging from this analysis, eight are associated with a given pathosystem or biological model (e.g. 'blight & rice', 'tomato & nematode', 'virus & viruses', etc.). The only two exceptions are clusters associated with the keywords 'strategies & management' and 'defense & immunity'.

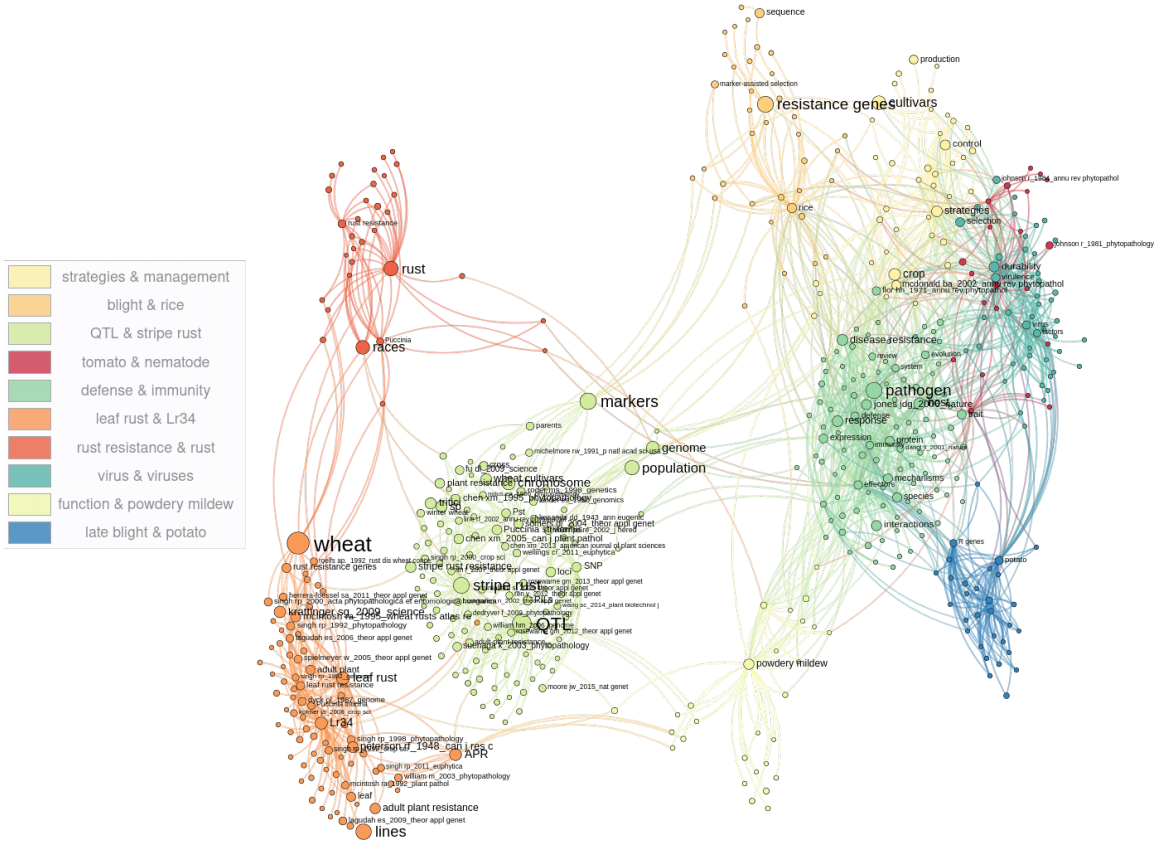


Fig. 6: Network and clustering of the main keywords and co-cited articles. This representation was drawn with the Network Mapping tool on CorTexT Manager, based on the list of 100 extracted keywords and the 'Cited References' field, with 500 nodes and all other parameters set to default parameters.

3.4 Without the model organism's layer, four clusters emerge

To lower the impact of organisms keywords on clustering analyses, a second keyword extraction was performed and refined manually. The first list of all WOS keywords referenced in the articles contained 4334 unique keywords. We kept only keywords with more than 5 occurrences in the corpus. Then we refined manually the list of 513 remaining elements to keep only keywords neither too general nor too specific nor referring to an organism name. We thus obtained a final list of 62 keywords listed in Tab. 1. The same clustering based on co-occurring keywords and co-cited articles was performed, with 150 nodes represented (Fig. 7).

This analysis highlights four main clusters, based on a network of keywords and co-cited articles. They gather articles around:

1. Identification and selection of genetic determinisms (main keywords: 'QTL', 'marker-assisted-selection', 'R-gene', 'locus', and 'GWAS'), hereafter referred to as the *'QTL'*

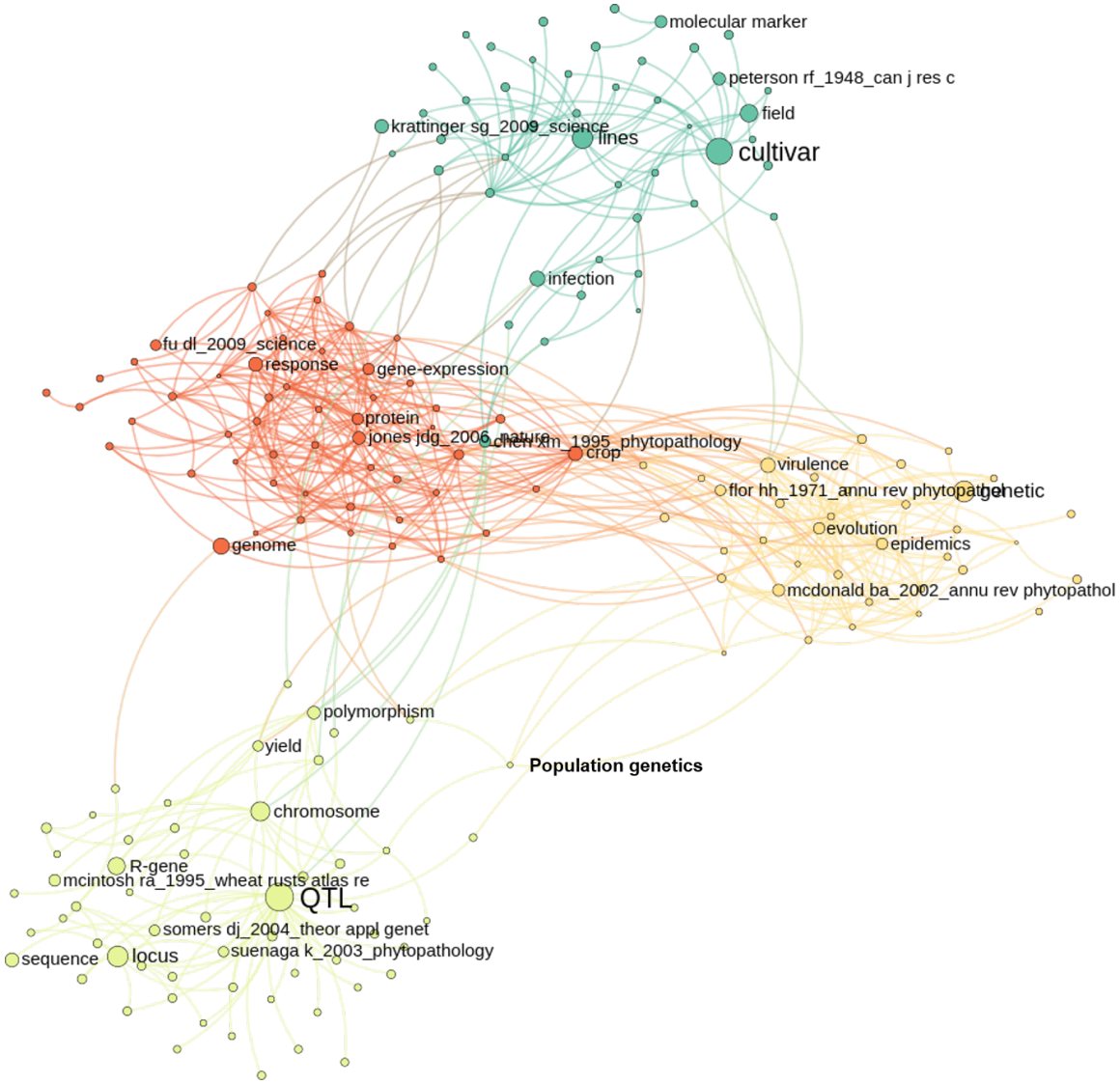


Fig. 7: Network and clustering of the refined list of keywords and the co-cited articles. This representation was drawn with the Network Mapping tool on CorTexT Manager platform, based on the refined list of 62 keywords and the 'Cited References' field, with 150 nodes and all other parameters set to default parameters. The 'population genetics' keyword has been put in bold to highlight its position in the network.

cluster,

2. Evaluation and characterisation of pathogenicity (main keywords: 'lines', 'cultivar', 'field' and 'infection'), hereafter referred to as the *'Field evaluation'* cluster.

3. Molecular investigations into plant-pathogen interactions (main keywords: 'protein', 'response', 'gene-expression', 'effector' and 'genome'), hereafter referred to as the *'molecular interactions'* cluster,

4. Epidemiology of plant diseases and evolution of pathogen populations, (main keywords: 'evolution', 'epidemics', 'fitness', 'virulence', and 'simulation model'), hereafter referred to as the *'epidemiology and evolution'* cluster,

The first and second clusters are focused mainly on research studies upstream of overcoming resistance, with selection of resistant plants, or their characterisation in the field, respectively. They group into the *'plant breeding'* cluster in the Opinion Note. Conversely, the third and fourth clusters are dedicated to the study downstream of overcoming resistance, with the understanding of mechanisms and consequences of such losses of resistance efficacy.

3.5 'Population genetics' is not found as a link among clusters

In the latter network, the 'population genetics' keyword is placed between the *'epidemiology and evolution'* cluster and the *'QTL'* cluster. Few papers (48 out of 1783, Tab. 1) refer to population genetics, and most of these studies employ population genetics as a specific and simplified tool applied to empirical work, and not as a conceptual discipline.

Among these articles, we found two kinds of studies:

First, some articles refer to population genetics only as a prerequisite for QTL and GWAS analyses. Indeed, association analyses require prior knowledge of population structures to perform accurate investigations on genetic determinisms. In these cases, population genetics relates to plants and not pathogens. The proximity of the 'population genetics' keyword to the *'QTL'* cluster reflects this methodological consideration.

Second, other articles referring to population genetics investigate pathogen population structures using neutral molecular markers to study resistance durability. These articles reflect the main body of population genetic studies, which aim to infer key elements of pathogen biology from neutral population structure analyses (e.g. the extent of gene flow as a proxy for migration rate or the variation of within-population indices as an indicator

Table 1: Refined Table of 62 keywords, with their number of occurrence in distinct documents, and their total number of occurrence in the corpus.

Keyword	**Number of distinct documents**	**Number of occurrences**
linkage	1311	3701
hypersensitive response	1050	2574
R-gene	967	2659
genetic	836	1636
virulence	789	3048
oxidative	776	1543
breeding	769	1499
cultivar	742	1713
crop	602	1217
lines	556	1244
field	516	878
locus	469	1036
infection	457	813
response	409	695
sequence	396	626
genome	386	895
QTL	333	1814
protein	327	904
chromosome	321	761
varieties	304	628
RNA	302	730
expression	300	577
yield	279	444
management	212	356
evolution	212	443
polymorphism	207	318
molecular marker	176	271
marker-assisted selection	172	226
DNA	154	250
functional	149	213
PCR	140	187
epidemics	139	257
effector	132	402
susceptibility	132	238
cloning	116	157
inheritance	108	164
mutation	107	188
GMO	100	162
LRR	75	145
receptor	74	158
fitness	70	150
genetic diversity	65	88
simulation model	62	92
transcriptome	61	100
necrosis	59	116
kinase	52	101
transcription factor	52	87
heritability	52	78
genome-wide association	51	144
nucleotide-binding	49	91
population genetics	48	76
recombination	46	75
cell-wall	45	72
aggressiveness	42	88
mixture	38	72
secretion	37	66
in-vitro	35	49
dispersal	31	43
transporter	28	52
signaling pathways	17	24
molecular characterization	13	15
local adaptation	3	6

of reproductive mode). Knowledge of these key biological aspects is of paramount importance to feed into epidemiological models and, ultimately, to forecast the risk of resistance breakdown (Hessenauer et al., 2021; Milgroom and Peever, 2003). An important node of the *'epidemiology and evolution'* cluster is the review by McDonald and Linde (2002), which considers population genetics to draw a typology of risk of resistance breakdown depending on pathogen features. Overall, these 'neutral' population genetic studies disregard selection as an evolutionary force. Population genomics would help to disentangle signatures of selection along pathogen genomes hence providing an opportunity to indirectly assess resistance durability (Plissonneau et al., 2017; Stukenbrock and Bataillon, 2012). However, such studies applied to plant pathogens are still underrepresented (Möller and Stukenbrock, 2017). In this respect, within the corpus of plant resistance durability literature, only 2 of the 1783 articles were assigned to the 'population genomics' keyword. Conversely, major studies on population genomics in plant pathogens are missing because they do not refer explicitly to resistance durability (e.g De Jonge et al., 2012; Hartmann et al., 2017; Stukenbrock et al., 2011)

Besides these two kinds of studies referring to population genetics, some articles of the corpus study the genetic variability of genes under selection (e.g. those encoding effectors) in pathogen populations without being assigned to the 'population genetics' keyword. Although these articles do not necessarily highlight the terms 'population genetics', their analyses are based on classical population genetics tools and concepts. Such empirical studies of genetic evolution are more often linked to the *'molecular interactions'* cluster, and do not associate genetic structures at selected loci to neutral genetic diversity during the overcoming of a resistance (Plissonneau et al., 2017).

3.6 Clusters differ in the temporal evolution of their associated keywords

The distribution of the number of articles in the corpus through time highlights a growing interest in the research field of plant resistance durability, from the founding paper of Johnson (1981) (Fig. 1).

To understand the evolution of the research field on plant resistance durability, a temporal representation of the clustering based on the list of 62 manually refined keywords was performed (Fig. 8). This figure represents the main keywords associated with a temporal partition of each cluster. Five homogeneous and overlapping time periods were defined, each time period containing the same number of articles. Fig. 8 shows a very stable cluster based on the 'QTL' keyword, which would correspond to the *'QTL'* cluster previously

identified without the temporal inclusion (Fig. 7). The *'epidemiology and evolution'* cluster is also evenly represented through time, with only slightly different representative keywords. For this cluster, a recent change in the approach is highlighted by the switch from the 'virulence' keyword to the 'aggressiveness' keyword, reflecting the increase in interest in quantitative resistances. The remaining studies show a large reshuffling of keywords through time, with rapidly evolving clusters, especially since 2010. Clusters are difficult to disentangle due to successive splitting and merging events of keywords associated with molecular studies. This may reflect a turn-over of methods in a constantly evolving research field.

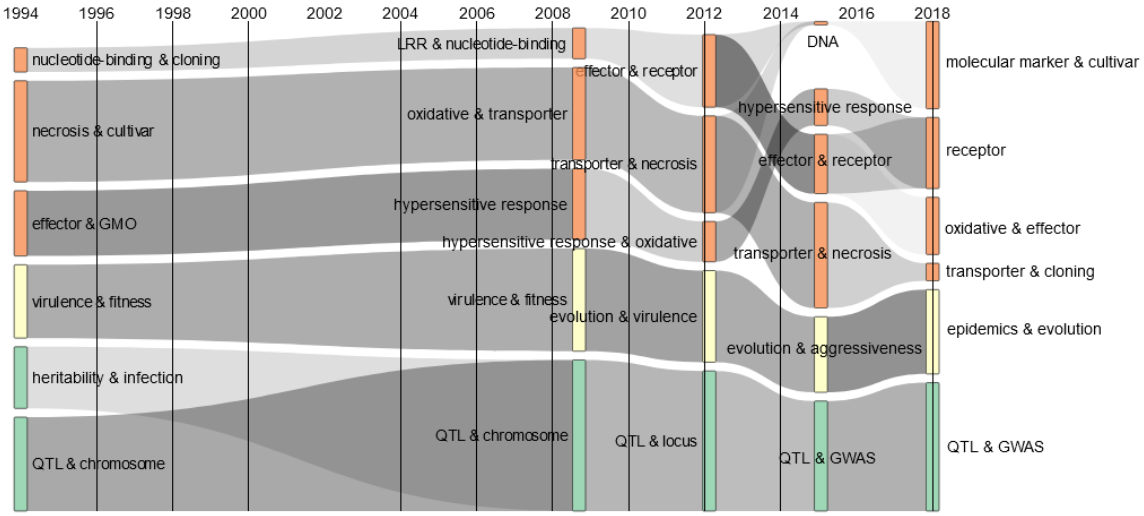


Fig. 8: Temporal evolution of the clustering, based on the refined list of keywords and the co-cited articles. This representation was drawn with the temporal Network Mapping tool on CorTexT Manager, based on the refined list of 62 keywords and the 'Cited References' field, with 150 nodes and all other parameters set to default parameters. Five time periods were selected, overlapping and homogeneous, so that each time period contained the same number of articles.

References

Cowger, C., & Brown, J. K. (2019). Durability of Quantitative Resistance in Crops: Greater Than We Know? *Annual Review of Phytopathology*, *57*, 253–277.

De Jonge, R., Van Esse, H. P., Maruthachalam, K., Bolton, M. D., Santhanam, P., Saber, M. K., Zhang, Z., Usami, T., Lievens, B., Subbarao, K. V., & Thomma, B. P. (2012). Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(13), 5110–5115.

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME Journal*, *11*(5), 1189–1204.

Hessenauer, P., Feau, N., Gill, U., Schwessinger, B., Brar, G. S., & Hamelin, R. C. (2021). Evolution and Adaptation of Forest and Crop Pathogens in the Anthropocene. *Phytopathology®*, *111*(1), 49–67.

Johnson, R. (1981). Durable resistance: definition of, genetic control, and attainment in plant breeding. *Phytopathology*, *71*(6), 567.

McDonald, B. A., & Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, *40*, 349–379.

Milgroom, M. G., & Peever, T. L. (2003). Population biology of plant pathogens: the synthesis of plant disease epidemiology and population genetics. *Plant Disease*, *87*(6), 608–617.

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, *15*(12), 756–771.

Plissonneau, C., Blaise, F., Ollivier, B., Leflon, M., Carpezat, J., Rouxel, T., & Balesdent, M.-H. (2017). Unusual evolutionary mechanisms to escape effector-triggered immunity in the fungal phytopathogen ¡i¿Leptosphaeria maculans¡/i¿. *Molecular Ecology*, *26*(7), 2183–2198.

Stukenbrock, E. H., & Bataillon, T. (2012). A Population Genomics Perspective on the Emergence and Adaptation of New Plant Pathogens in Agro-Ecosystems (J. Heitman, Ed.). *PLoS Pathogens*, *8*(9), e1002893.

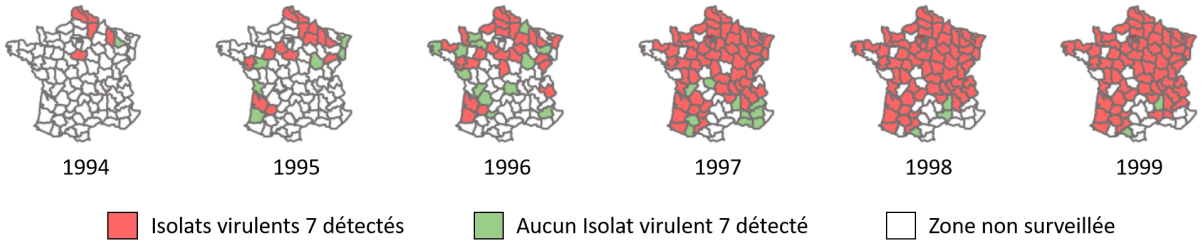
Stukenbrock, E. H., Bataillon, T., Dutheil, J. Y., Hansen, T. T., Li, R., Zala, M., McDonald, B. A., Wang, J., & Schierup, M. H. (2011). The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen Mycosphaerella graminicola and its wild sister species. *Genome Research*, *21*(12), 2157–2166.

1.3 Cas d'étude : le contournement de la résistance RMlp7 par *Melampsora larici-populina*

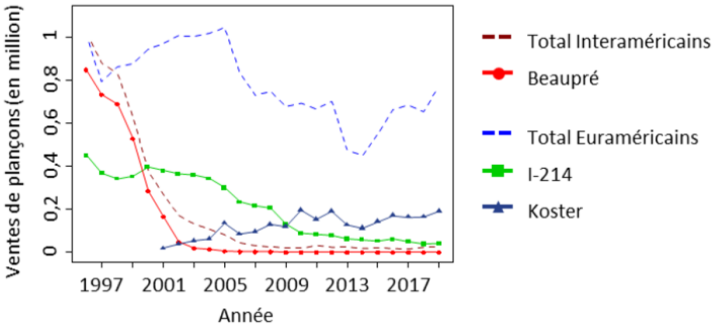
Le pathosystème peuplier-*Melampsora larici-populina* étudié dans cette thèse a fait l'objet d'études dans chacun des trois domaines de recherches préalablement décrits (le domaine de la sélection variétale, le domaine de l'étude des interactions moléculaires plantes-pathogènes et le domaine de l'épidémiologie et de l'évolution). Pour lutter contre cet agent pathogène, des efforts considérables de sélection variétale ont été réalisés sur les peupliers (PINON *et al.*, 1998; BRESSON *et al.*, 2011), des études d'interactions moléculaires ont aidé à mieux comprendre les mécanismes de résistance impliqués (HACQUARD *et al.*, 2011; PERSOONS *et al.*, 2014; PETRE *et al.*, 2016; NVSVROT *et al.*, 2020), et des études épidémiologiques et évolutives ont pu mettre en lumière les conséquences des contournements (BARRÈS *et al.*, 2008; XHAARD *et al.*, 2011; PERSOONS *et al.*, 2017).

Entre les années 1970 et 2020, l'amélioration variétale du peuplier a permis de sélectionner plusieurs résistances qualitatives pour limiter les pertes économiques occasionnées par Mlp, notamment huit qui ont été déployées (RMlp1 à RMlp8). Cependant, les plantations monoclonales ont rendu les peupliers particulièrement vulnérables à la rouille (GÉRARD *et al.*, 2006). Ces huit résistances ont toutes été contournées par Mlp (PINON ET FREY, 2005; FABRE *et al.*, 2021a). Le contournement le plus dommageable qui ait été documenté est celui de la résistante RMlp7. Cette résistance est portée notamment par le cultivar Beaupré qui a été planté très largement dans le Nord de la France entre 1982 et 1996. En 1996, ce cultivar représentait ainsi 57% des plants vendus en France, et jusqu'à 80% des plants vendus dans le nord de la France. Un évènement adaptatif majeur a eu lieu en 1994 lorsque les premiers isolats de Mlp porteurs de la virulence 7 ont émergé, menant à un contournement de la résistance RMlp7 à partir de foyers situés en Belgique et dans le nord de la France. La virulence 7 s'est alors répandue dans les populations de Mlp de l'ensemble du territoire français en moins de 5 ans (FIGURE 1.4a), entraînant des dégâts économiques considérables (XHAARD *et al.*, 2011; PERSOONS *et al.*, 2017). Suite à ce contournement de résistance et après avoir dominé les plantations pendant plus de 20 ans, les plantations du cultivar Beaupré ont largement diminué, jusqu'à un arrêt total des ventes en 2005 (FIGURE 1.4b).

Le peuplier est un hôte pérenne et les plants de Beaupré n'ont pas été retirés lors du contournement, ce qui a entraîné une pression de sélection forte et durable sur les populations de Mlp. Le contournement de la résistance RMlp7 a ainsi fortement impacté la structure des populations de Mlp (XHAARD *et al.*, 2011; PERSOONS *et al.*, 2014, 2017, 2022). PERSOONS *et al.*, 2017 ont étudié l'évolution de la structure génétique des populations de Mlp sur le territoire français. Ils se sont appuyés pour cela sur les profils de virulence et des marqueurs neutres appliqués à une collection d'individus prélevés entre 1982 et



(a) Détection des isolats virulent 7 chez Mlp



(b) Évolution des ventes de peupliers

FIGURE 1.4 – Contournement de la résistance RMlp7 par Mlp. (a) Suivi spatiotemporel de la dispersion des isolats Mlp virulents 7 sur le territoire français (d'après FABRE *et al.*, 2021a). (b) Évolution des ventes de trois cultivars de peupliers (Beaupré, I-214 et Koster) et des types de croisements (Interaméricains et Euraméricains) suite au contournement de la résistance RMlp7 (données du Ministère de l'Agriculture et de l'Alimentation, d'après FABRE *et al.*, 2021a).

2011, soit avant, pendant et après le contournement de la résistance RMlp7. Trois groupes génétiques distincts ont été mis en évidence (FIGURE 1.5) :

1. Un groupe ancestral qui a décliné à partir de 1994 puis disparu en 1998 (appelé groupe « Fossile »), composé uniquement d'individus avirulents 7.
2. Un groupe récent dont l'émergence en 1994 coïncide avec la détection du contournement de la résistance RMlp7 (appelé groupe « Cultivé »), composé presque exclusivement d'individus virulents 7.
3. Un groupe prédominant au sud-est de la France (XHAARD *et al.*, 2011) où les populations de Mlp se trouvent principalement sur des peupliers sauvages (appelé groupe « Sauvage »), composé majoritairement d'individus avirulents 7.

Le contournement de la résistance RMlp7 a ainsi entraîné l'émergence du groupe génétique « Cultivé » très homogène qui prédomine au nord de la France depuis plus de 20 ans (FABRE *et al.*, 2021a). Le phénotype de virulence ayant été soumis à une forte sélection, ce groupe génétique ne contient presque plus que des individus porteurs de la virulence 7.

Le génome de *M. larici-populina* est un génome de grande taille (110 Mb) avec une grande concentration en éléments transposables (50%) et plus de 16000 gènes, répartis

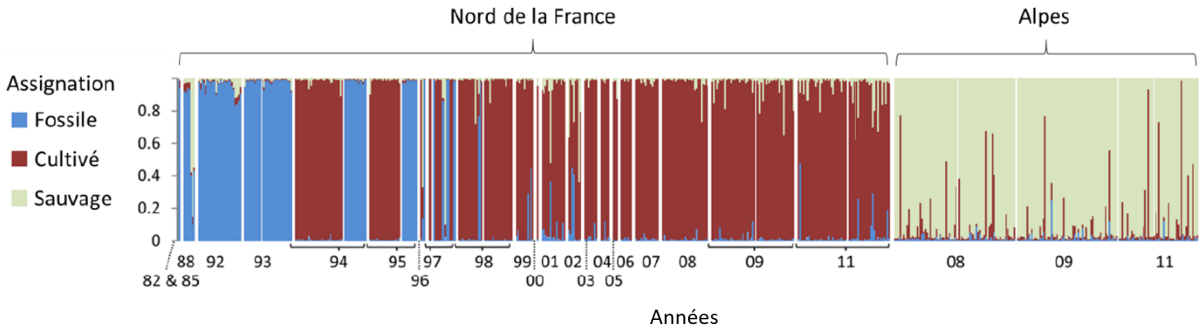


FIGURE 1.5 – Evolution de la structure génétique des populations de Mlp sur le territoire français (adapté d'après PERSOONS, 2015; PERSOONS *et al.*, 2017). Chaque barre verticale représente l'assignation Bayésienne d'un individu aux trois groupes génétiques identifiés (Fossile, Cultivé et Sauvage). Ces assignations représentent les probabilités d'appartenance aux différents groupes génétiques. La longueur de chaque segment est proportionnelle à la probabilité d'appartenance au groupe génétique correspondant. Les individus sont regroupés sur le graphique selon deux provenances géographiques distinctes : le nord de la France et les Alpes.

en 909 familles (DUPLESSIS *et al.*, 2011). Des études de génomique des populations ont rendu possible l'identification dans le génome de Mlp d'un gène d'avirulence candidat appelé *AvrMlp7* (PERSOONS *et al.*, 2022). Suite à cette identification et en marge de ma thèse, nous avons étudié le mécanisme de virulence impliqué en combinant des approches de génétique des populations et des approches moléculaires (Annexe, LOUET *et al.*, 2021). Cette étude s'est basée sur 281 individus de Mlp prélevés entre 1989 et 2016, soit avant, pendant, et après le contournement de la résistance RMlp7. Elle a permis de caractériser le phénotype de virulence et le génotype de chaque individu au locus candidat *AvrMlp7*. Cette caractérisation a mis en évidence deux mécanismes d'altération distincts au locus candidat : une mutation Single Nucleotide Polymorphism (SNP) non-synonyme (notée avr) et une large délétion au niveau de cette région génomique (notée Ø). Ces deux allèles peuvent être associés au sein des individus virulent. Ils sont statistiquement corrélés à la virulence des isolats vis-à-vis de la résistance RMlp7 (FIGURE 1.6). Grâce à la chronoséquence des individus au cours du contournement de la résistance RMlp7, cette étude a également montré une corrélation entre les trajectoires évolutives des altérations au locus candidat et le remplacement du groupe génétique « Fossile » par le groupe génétique « Cultivé ».

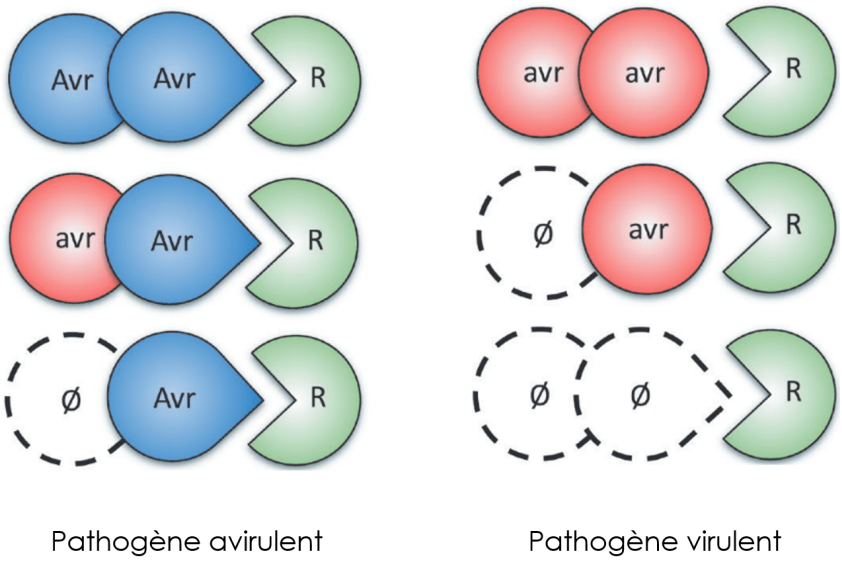


FIGURE 1.6 – Schéma du modèle gène-pour-gène appliqué au pathosystème peuplier-Mlp lors du contournement de la résistance RMlp7 (d'après LOUET *et al.*, 2021). Avr et avr représentent les produits des allèles avirulent et virulent de l'agent pathogène, respectivement. Ø représente l'absence de produit due à la délétion au niveau de la région génomique du locus candidat. R représente un récepteur immunitaire de l'hôte résistant, capable de reconnaître uniquement le produit Avr. En cas de détection d'au moins un produit Avr par le récepteur immunitaire de l'hôte, l'infection est stoppée et le pathogène est avirulent. En cas de non détection des produits avr ou Ø par le récepteur de l'hôte, l'infection a lieu et le pathogène est virulent.

1.4 Objectifs

Dans cette partie, je souhaite mettre en avant le lien entre les variations démographiques et l'évolution génétique d'une population d'agents pathogènes responsables d'une maladie émergente réelle, dans un système pour lequel la composante temporelle est déterminante. J'utilise pour cela l'apport conjoint de la modélisation et des données empiriques pour caractériser l'évolution génétique des populations de Mlp lors du contournement de la résistance RMlp7. Ce contournement est une situation où l'équilibre démographique de la population d'agents pathogène n'est pas respecté mais pour laquelle la pression de sélection est stable et unidirectionnelle. J'utilise ce système particulier pour construire un modèle démogénétique adapté et répondre aux objectifs suivants :

1. Déterminer les conditions d'existence d'un contournement de résistance par la modélisation, et notamment l'influence du niveau de ploïdie et du cycle de vie de l'agent pathogène.
2. Déterminer l'interaction entre les variations démographiques et la structure génétique neutre résultant d'un évènement de contournement de résistance.
3. Inférer le scénario de contournement ainsi que les paramètres écologiques et démo-

graphiques ayant régi le contournement de la résistance RMlp7 par l'analyse conjointe des données empiriques et du modèle.

Les chapitres 2, 3 et 4 répondent respectivement aux objectifs 1, 2 et 3 de cette partie. Deux études théoriques de modélisation sont présentées dans les chapitres 2 et 3. Les apports théoriques de ces chapitres seront appliqués à la compréhension du contournement de la résistance RMlp7 par Mlp. Le chapitre 4 fait la jonction entre la démogénétique théorique et les données empiriques en intégrant les données au modèle lors d'une analyse d'inférence par la méthode ABC.

Chapitre 2

Étude épidemio-évolutive d'un contournement de résistance

L'objectif de ce chapitre est de déterminer les conditions d'existence d'un contournement de résistance et d'identifier les déterminants de la durabilité d'une résistance. Je m'affranchis pour cela de certaines hypothèses de modélisation généralement admises, de façon à mieux respecter les spécificités de notre modèle biologique quant à la ploïdie et au cycle de vie. Je souhaite notamment comprendre comment les allèles virulents ont pu se répandre aussi rapidement dans la population de Mlp lors du contournement de la résistance RMlp7 alors qu'ils sont restés latents plusieurs années dans la population précontournante (LOUET *et al.*, 2021).

Pour répondre à ces questions, deux approches de modélisation distinctes ont été développées :

1. Une modélisation par simulations stochastiques pour étudier les trajectoires évolutives et démographiques de la population d'agents pathogènes lors d'un contournement de résistance.
2. Une modélisation par résolution analytique déterministe pour étudier les principaux équilibres du système et leur stabilité.

Ces deux approches permettent de modéliser l'évolution d'une population d'agents pathogènes en considérant conjointement sa démographique et sa génétique au locus sous sélection.

2.1 Article 2. Impact of ploidy and pathogen life cycle on resistance durability

Une première approche par simulations stochastiques nous a permis d'évaluer l'influence d'un ensemble de paramètres d'intérêt sur la probabilité d'aboutir à un contournement et sur la dynamique de fixation de la virulence, deux mesures de la durabilité d'une résistance qui ne sont pas nécessairement corrélées (VAN DEN BOSCH ET GILLIGAN, 2003).

Les paramètres étudiés se divisent en plusieurs catégories : 1) Deux paramètres principaux, intrinsèquement liés au cycle de vie de l'agent pathogène considéré : le niveau de ploïdie de l'agent pathogène (haploïde ou diploïde) et l'alternance ou non de l'agent pathogène sur une autre espèce d'hôte lors de sa reproduction sexuée ; 2) Deux paramètres démographiques de la population d'agents pathogènes : le taux de croissance et le taux de migration ; 3) Un paramètre génétique de la population d'agents pathogènes : la quantité de variabilité génétique préexistante au locus d'avirulence ; 4) Un paramètre de gestion humaine : la proportion d'hôtes résistants déployés dans le paysage agricole.

Ces résultats sont ici présentés sous la forme d'un article de recherche intitulé « Impact of ploidy and pathogen life cycle on resistance durability », qui a été publié en novembre 2021 dans le premier volume de la revue PEER COMMUNITY JOURNAL.

Peer Community Journal

Section: Evolutionary Biology

RESEARCH ARTICLE

Published
2021-11-23

Cite as
Méline Saubin, Stéphane De Mita, Xujia Zhu, Bruno Sudret and Fabien Halkett (2021) *Impact of ploidy and pathogen life cycle on resistance durability*, Peer Community Journal, 1: e8.

Correspondence
fabien.halkett@inrae.fr

Peer-review
Peer reviewed and recommended by PCI Evolutionary Biology, https://doi.org/10.24072/pci.evolbiol.100131

CC BY

This article is licensed under the Creative Commons Attribution 4.0 License.

Impact of ploidy and pathogen life cycle on resistance durability

Méline Saubin [1], Stéphane De Mita [2], Xujia Zhu [3], Bruno Sudret [3], and Fabien Halkett [1]

Volume **1** (2021), article e8

https://doi.org/10.24072/pcjournal.10

Abstract

The breeding of resistant hosts based on the gene-for-gene interaction is crucial to address epidemics of plant pathogens in agroecosystems. Resistant host deployment strategies are developed and studied worldwide to decrease the probability of resistance breakdown and increase the resistance durability in various pathosystems. A major component of deployment strategies is the proportion of resistant hosts in the landscape. However, the impact of this proportion on resistance durability remains unclear for diploid pathogens with complex life cycles. In this study, we modelled pathogen population dynamics and genetic evolution at the virulence locus to assess the impact of the ploidy (haploid or diploid) and the pathogen's life cycle (with or without host alternation) on resistance durability. Ploidy has a strong impact on evolutionary trajectories, with much greater stochasticity and delayed times of resistance breakdown for diploids. This result emphasises the importance of genetic drift in this system: as the virulent allele is recessive, positive selection on resistant hosts only applies to homozygous (virulent) individuals, which may lead to population collapse at low frequencies of the virulent allele. We also observed differences in the effect of host deployment depending on the pathogen's life cycle. With host alternation, the probability that the pathogen population collapses strongly increases with the proportion of resistant hosts in the landscape. Therefore, resistance breakdown events occurring at high proportions of resistant hosts frequently amount to evolutionary rescue. Last, life cycles correspond to two selection regimes: without host alternation (soft selection) the resistance breakdown is mainly driven by the migration rate. Conversely, host alternation (hard selection) resembles an all-or-nothing game, with stochastic trajectories caused by the recurrent allele redistributions on the alternate host.

[1]Université de Lorraine, INRAE, IAM, F-54000 Nancy, France, [2]INRAE, Cirad, IRD, Montpellier SupAgro, Université de Montpellier, PHIM, F-34000 Montpellier, France, [3]Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Stefano-Franscini-Platz 5, CH-8093, Switzerland

Contents

1 Introduction . . . 2
2 Model description . . . 4
2.1 Model overview . . . 4
2.2 Demographic evolution of the pathogen population . . . 6
2.3 Genetic evolution of the pathogen population . . . 7
2.4 Implementation of model analyses . . . 7
3 Results . . . 10
3.1 Model behaviour . . . 10
3.2 Sensitivity analyses . . . 10
3.3 Patterns of virulence fixation . . . 10
3.4 Variations in the speed of virulence spread . . . 12
4 Discussion . . . 14
4.1 Deep impact of the ploidy on resistance durability . . . 14
4.2 Life cycles impose different selection regimes and lead to contrasted evolutionary trajectories . . . 17
5 Conclusion . . . 18
6 Perspectives . . . 18
7 Data accessibility . . . 19
8 Acknowledgements . . . 19
9 Funding . . . 19
10 Conflict of interest disclosure . . . 19
Annex A: Sobol indices . . . 20
Annex B: Supplementary figures . . . 21
References . . . 26

1. Introduction

Plant pathogens can quickly evolve (Perkins et al., 2013), and the loss of host genetic diversity in agroecosystems compared to natural ecosystems can enhance the spread of epidemics (Burdon and Thrall, 2008; Garrett et al., 2009; Haas et al., 2011; Mundt, 2002; Ostfeld and Keesing, 2012; Zhan et al., 2015). In this context, many plant protection strategies are developed and studied worldwide (Bousset and Chèvre, 2013), particularly spatio-temporal host resistance deployment strategies (Bousset et al., 2018; Burdon et al., 2014; Djian-Caporalino et al., 2014; Fabre et al., 2015; Gilligan and Bosch, 2008; Mundt, 2002; Rimbaud et al., 2018b; Sapoukhina et al., 2009). However these modelling studies seldom account for pathogen differences in life cycle and ploidy levels.

While quantitative resistance has gained interest (Pilet-Nayel et al., 2017), the breeding of disease resistant plants is still often based on the gene-for-gene interaction (Person et al., 1962; Zhan et al., 2015). In the simplest case of specific response, the result of the infection is determined by the interaction between a locus in the plant (the resistance gene) and in the pathogen (the avirulence gene) (Flor, 1971). This interaction leads to an all-or-nothing response and therefore such resistances are called qualitative. Qualitative resistances often rely on the recognition of a specific pathogen molecule (an effector protein for instance) by a plant immune receptor (Lo Presti et al., 2015). If the pathogen is recognised by the plant, the infection is stopped and the plant is called *resistant*. But the pathogen species evolves in multiple ways to escape host recognition (Rouxel and Balesdent, 2017). When a pathogen can infect a resistant host it is called *virulent*, as opposed to *avirulent* individuals. For avirulent individuals, if the product of the avirulence gene is not recognised by the plant, the infection occurs and the plant is called *susceptible*.

Hence, virulent individuals can infect both susceptible and resistant hosts, while avirulent individuals can only infect susceptible hosts. In its simplest cases, the avirulence gene exists in two versions: the avirulent *Avr* allele and the virulent *avr* allele. The plant resistance can thus be overcome by a mutation of the *Avr* allele which modifies the pathogen recognition by the plant. The *Avr* allele is then replaced by a virulent *avr* allele which leads to a virulent pathogen (Stukenbrock and McDonald, 2009a).

In natural systems, the constant turnover of resistance and avirulence genes results from a strong coevolutionary interaction between both species (Zhan et al., 2014), represented by the concept of arms-race (Brown and Tellier, 2011). On both sides, the most adapted allele can spread in the population, sometimes replacing alleles conferring lower fitness to individuals (Brown and Tellier, 2011; Persoons et al., 2017). These genes are under strong selective pressure and at each selective event a selective sweep can occur and drastically reduce the genetic diversity of both species (Oleksyk et al., 2010; Terauchi and Yoshida, 2010). In natural populations, rare host genotypes can be maintained by negative frequency-dependent selection, resulting in the preservation of host polymorphism (Lewontin, 1958). In agroecosystems, however, pure crops of resistant hosts hinder this maintenance of polymorphism (Zhan et al., 2015). Therefore, the issue of such resistance deployments is often a resistance breakdown, *i.e.* the failing of the host to remain resistant to the pathogen, which can result in severe epidemics (Brown and Tellier, 2011; Burdon et al., 2016; Johnson, 1984; Pink and Puddephat, 1999). Such a resistance breakdown can occur more or less quickly, depending on the pathosystem considered and the environmental conditions (Van den Bosch and Gilligan, 2003). This observation raises the question of resistance durability, which can be defined as the time until the virulent population reaches a given threshold in the pathogen population. Definitions of resistance durability can have diverse acceptations depending on the threshold considered (Brown, 2015; Carolan et al., 2017; Lof et al., 2017; Pacilly et al., 2018; Pietravalle et al., 2006; Rimbaud et al., 2021; Van den Bosch and Gilligan, 2003). Considering several thresholds can help in capturing different steps of the pathogen dynamics.

Resistance durability becomes a major economical issue when epidemics impact crop yields. Therefore, it has often been studied through the modelling of epidemics spread in agricultural landscapes (Rimbaud et al., 2021). Such models can couple epidemiological and evolutionary processes, and often aim to study the influence of different biological parameters on the emergence of pathogens, their specialisation to the host plant, the evolutionary dynamics of virulence, or on the resistance durability (Van den Bosch and Gilligan, 2003). *Virulence* is defined here as the ability for a pathogen individual to infect a resistant host, in accordance to the phytopathology literature (Flor, 1971; McDonald and Linde, 2002). These parameters can be specific to the host plant (proportion of resistant host in the landscape, their spatial and temporal distribution) or to the pathogen (life cycle, mutation rate, dispersal) (Fabre et al., 2012, 2015; Papaïx et al., 2015, 2018; Soularue et al., 2017). These models often represent haploid pathogens with a virulent and an avirulent genotype, evolving purely asexually on a landscape composed of compartments, gathering resistant or susceptible hosts (Lof et al., 2017; Lof and Werf, 2017; Pacilly et al., 2018; Pietravalle et al., 2006). Regarding the pathogen, high risks of resistance breakdown are observed for pathogen populations with high gene flow and mutation rates, large effective population sizes, and partially asexual reproductions (McDonald and Linde, 2002). Regarding the host, the increase in the proportion of resistant hosts should increase the selection pressure, hence weakening the resistance durability (Pietravalle et al., 2006; Van den Bosch and Gilligan, 2003). However, a large proportion of resistant hosts also reduces the initial size of the pathogen population and thus the risk of resistance breakdown (Pacilly et al., 2018), partly because a small population size reduces the likelihood that a virulent individual will emerge through mutation.

However, the impact of host resistance deployment on resistance durability remains unclear when the pathogen is diploid (like rust fungi, oomycetes, or nematodes). When the product of the avirulence gene is a specific molecule like an effector protein, the pathogen is virulent only if this product is not detected by the product of the corresponding resistance gene in the host (Stukenbrock and McDonald, 2009a). Therefore, for a diploid individual the pathogen is virulent only if the products of both alleles avoid detection by the host. In other words, in the classical

gene-for-gene interaction the virulent allele is recessive (Thrall et al., 2016). Consequently, a heterozygous *Avr*/*avr* individual is phenotypically avirulent, and the selective advantage of the virulence is effective among homozygous *avr*/*avr* individuals only. At low frequency, *avr* alleles are then carried by heterozygous individuals and mostly subjected to drift.

Diploid pathogens exhibit a large variability of life cycles (Agrios, 2005). We can especially distinguish autoecious pathogens, which complete their life cycle on a unique host species, from heteroecious pathogens which need two different and successive host species to complete their life cycle (Lorrain et al., 2019; Moran, 1992). This presence or absence of an alternate host species could also affect the influence of host deployment strategy on resistance durability. Moreover, most studies focus on purely asexual pathogens, but the highest risks of resistance breakdown were observed for mixed reproduction systems (McDonald and Linde, 2002), with the best invaders combining high rates of asexual reproduction and rare events of sex (Bazin et al., 2014). Yet, the allelic redistribution resulting from a sexual reproduction event could have an even stronger impact on resistance durability when the pathogen is diploid.

To study resistance durability and evolutionary forces shaping the system, the understanding of the evolution of gene and virulence allele frequencies is needed. Coupling epidemiology and population genetics provides insights on both short and long time scales. It allows in particular detailed analyses of transition periods (Bolker et al., 2010; Day and Gandon, 2007; Day and Proulx, 2004), through variables like the pathogen population size, affecting both the disease incidence in epidemiology and the impact of genetic drift in population genetics (McDonald, 2004). This approach is also crucial for highlighting transient effects like evolutionary rescue, *i.e.* the genetic adaptation of a population to a new environment, thus preventing its extinction (Alexander et al., 2014; Martin et al., 2013). Evolutionary rescue as a process leading to resistance breakdown has not received consideration so far.

The virulence of pathogens can be associated with a fitness cost on susceptible hosts (Bousset et al., 2018; Laine and Barrès, 2013; Leach et al., 2001; Montarry et al., 2010; Nilusmas et al., 2020; Thrall and Burdon, 2003), sometimes referred to as the cost of pathogenicity (Sacristán and García-Arenal, 2008). This fitness cost has been shown to have a strong impact on resistance durability (Fabre et al., 2012). However, depending on the avirulence gene considered, such a fitness cost is not systematic (see Leach et al., 2001 for a review). Therefore, in the absence of data, it could be more conservative of the risk of breakdown not to consider fitness cost while modelling resistance durability.

In this paper, we aim to broaden our understanding about the impact of the ploidy and the life cycle of pathogens on resistance durability. We used a non-spatialised model coupling population dynamics and population genetics to simulate the evolution of pathogens on susceptible and resistant hosts. We investigated the effects of resistant host deployment and pathogen demography on resistance durability, for a population of diploid and partially clonal pathogens, and compared the results to those obtained for haploid pathogens. Two different life cycles were implemented specifically: with or without host alternation for the sexual reproduction of the pathogen population. We assessed the resistance durability in two steps. First we examined the dynamics of fixation of the virulent allele in the population, and considered in parallel the cases when the pathogen population could go extinct, to highlight evolutionary rescue events. Then we focused on the invasion and resistance breakdown events, and disentangle the relationship between the durability of resistance and the dynamics of virulent populations after the invasion of the resistant plants.

2. Model description

2.1. Model overview.

The model is individual-based, forward-time and non-spatialised, and couples population dynamics and population genetics to study the evolution of a population of pathogens for a succession of generations. It allows us to follow the evolutionary trajectory of different genotypes at the avirulence locus of pathogens through time. We consider a life cycle usually found in temperate pathogen species, which alternate rounds of clonal reproduction with an annual event of

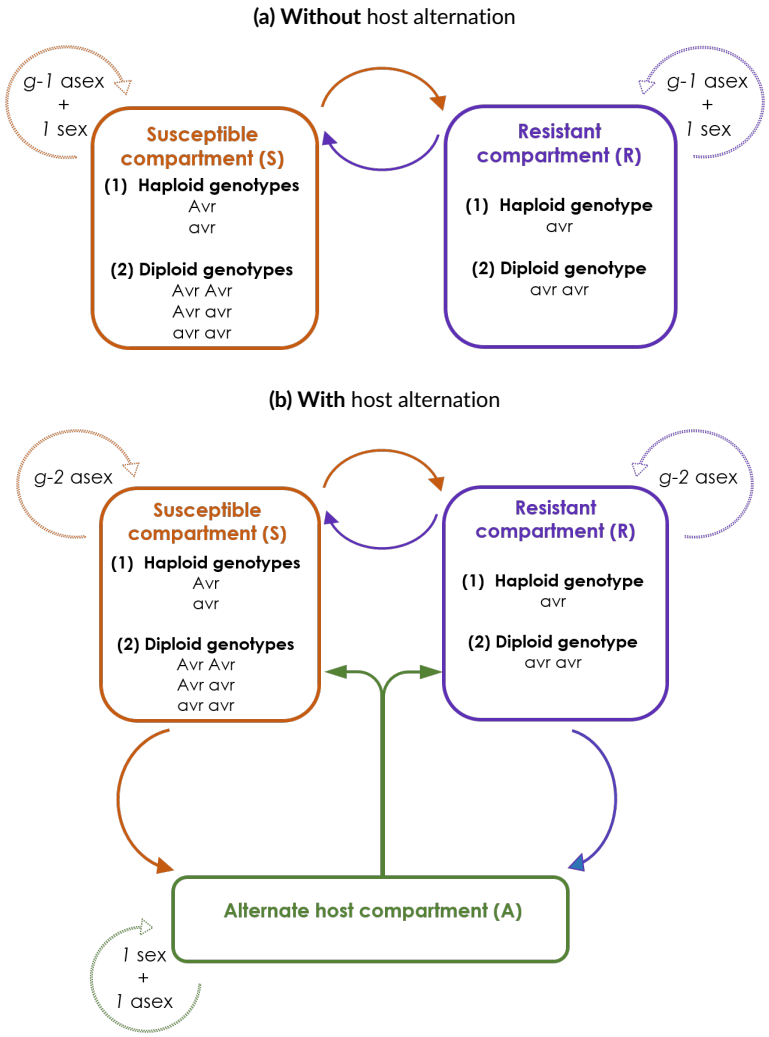


Figure 1 – Model representation for two different life cycles: (a) without or (b) with host alternation. *g* corresponds to the total number of generations (asexual plus sexual) in a year. Dashed arrows represent reproduction events, and solid arrows represent migration events occurring at each generation. *asex* stands for asexual reproduction events, and *sex* for sexual reproduction events. *avr* denotes the virulent recessive allele, and *Avr* the avirulent dominant allele.

sexual reproduction (Agrios, 2005). This model is designed in four variants to represent haploid or diploid pathogens with two distinct life cycles: with or without host alternation for the sexual reproduction (Figure 1). Without host alternation, the model represents the evolution in time of a population of pathogens on two static host compartments: a resistant compartment (R) and a susceptible compartment (S). Fixed carrying capacities of pathogens, K_R and K_S, are respectively assigned to R and S compartments and represent the maximum amount of pathogens that each compartment can contain. With host alternation, the alternate host compartment (A) is added, where the sexual reproduction takes place. This static compartment is assumed to be sufficiently large and thus with unbounded population size. Note that the life cycle with host alternation for haploid pathogens was added for the sake of comparison but has no real biological meaning, because no haploid pathogen display this life cycle.

2.2. Demographic evolution of the pathogen population.

2.2.1. Reproduction events. Each discrete generation corresponds to a reproduction event, either sexual or asexual. Each year is composed of g non-overlapping generations, with one annual sexual reproduction event followed by a succession of $g-1$ asexual reproduction events. In our simulations, we considered a year composed of $g = 11$ generations. At each reproduction event, parents give way to offspring and the new population is composed exclusively of new individuals. The within-compartment dynamics of the pathogen population are provided by the following equations:

Before each sexual reproduction event, a proportion *Reduct* of pathogen individuals is picked randomly to form the parental population. We fixed $Reduct = 0.2$ to cope to pathogen life cycles displaying drastic reduction in population size during sexual reproduction which usually takes place in winter.

For the sexual reproduction event itself, the population size is considered constant before and after the reproduction event:

$$N_{n+1} = N_n \quad (1)$$

With N_{n+1} the population size at generation $n+1$ and N_n the population size at generation n. Sexual offspring genotypes are obtained through random mating within the parental population.

For the asexual reproduction following the sexual reproduction in the A compartment in the life cycle with host alternation, the population growth is exponential, with the following relation:

$$N_{n+1} = r \times N_n \quad (2)$$

With r the growth rate of the pathogen population.

For each asexual reproduction in the R or S compartments, the population growth is logistic, with the following relation:

$$N_{n+1} = N_n + (r-1) \times N_n \times (1 - \frac{N_n}{K}) \quad (3)$$

With K the carrying capacity of the compartment (K_R or K_S for R or S compartment respectively). For all clonal reproduction events, offspring genotypes are drawn randomly from the parental population, with replacement, considering the same reproduction rate for all pathogen genotypes.

2.2.2. Migration events. A regular two-way migration event takes place each generation before the reproduction event, between individuals evolving in the R and S compartments. The number of migrant individuals is determined by a migration rate (*mig*) multiplied by the number of individuals on the compartment of origin. Migrant individuals succeed to invade the compartment of arrival, even if the number of individuals on this compartment reached the maximum carrying capacity. The population size on each compartment is restricted to the carrying capacity during reproduction events only, and not during migration events. Thereby, this choice enables the immigration of new pathogens regardless of the size of the population, as it is observed in natural populations for plant pathogens.

For the life cycle with host alternation, the annual sexual reproduction event coincides with the obligate migration of the entire pathogen population to and from the alternate host. The first migration event takes place once every year after $g-2$ asexual reproduction events in the R and S compartments (Figure 1). For this migration event, an established proportion of individuals *Reduct* is picked randomly from R and S compartments to migrate to the A compartment. All remaining individuals die in the R and S compartments, because the sexual reproduction is mandatory to complete the life cycle. After the two reproduction events (sexual and asexual) in the A compartment, the second migration event redistributes randomly all individuals from the compartment A to R and S compartments, in proportion to the relative size of R and S compartments (Figure 1).

2.3. Genetic evolution of the pathogen population.

To better highlight the effect of drift among other evolutionary forces, we did not consider mutation, that is, there is no change by chance of allelic state. This amounts to study evolution of the pathogen population from standing genetic variation (Barrett et al., 2008). The avirulence gene exists at two possible states: the *Avr* allele and the *avr* allele. For haploid pathogens, the *Avr* allele leads to avirulent individuals surviving only in the S compartment (and in the A compartment in the case of host alternation), while the *avr* allele leads to virulent ones capable to survive on all compartments without any fitness cost (Brown, 2015; Leach et al., 2001). For diploid pathogens, *Avr* is dominant and *avr* is recessive. Thus, individuals with genotypes *Avr*/*Avr*, *Avr*/*avr* and *avr*/*avr* survive with equal fitness in the S and A compartments, while only individuals with the virulent genotype *avr*/*avr* survive in the R compartment. Every avirulent individual (*Avr* for haploids, and *Avr*/*Avr* or *Avr*/*avr* for diploids) migrating to the R compartment dies before any subsequent migration or reproduction event.

Besides the demographic evolution of pathogen populations, the model describes the evolution of allelic and genotypic frequencies through generations in each compartment. Reproduction events can change allelic and genotypic frequencies. In particular, the annual sexual reproduction is the only event responsible for allele reshuffling in diploid individuals. For haploid pathogens, as only one locus is studied, the sexual reproduction event amounts to asexual reproduction, with differences in the size of the offspring population only.

Resistance durability is evaluated at four steps representing different proportions of virulent individuals in the population: (1) the time of apparition of the first virulent individual on the R compartment; (2) the time of invasion of the R compartment (1‰ of the R compartment occupied); (3) the time of resistance breakdown (1% of the R compartment occupied); and (4) the time of fixation of the virulence (all individuals are virulent, i.e only *avr* alleles remain). The thresholds of 1‰ and 1% were arbitrarily fixed to correspond to (i) the establishment of a pathogen population on the R compartment for the invasion and (ii) the detection of the virulent population on the R compartment for the resistance breakdown, respectively.

2.4. Implementation of model analyses.

The model was implemented in Python (version 3.7, Rossum, 1995), with the package "simuPOP" (Peng and Kimmel, 2005). The starting point of each replicate simulated was a population of 2000 individuals in the susceptible compartment. A proportion f_{avr} of virulent alleles was introduced initially in the pathogen population at Hardy-Weinberg equilibrium, as standing genetic variation. For diploid individuals, homozygous *avr*/*avr* individuals could therefore be initially present, depending on f_{avr}. All simulations were run with a fixed total carrying capacity for the host population size, $K = K_R + K_S = 100\,000$, but a variable proportion of the size of the R compartment $propR = \frac{K_R}{K}$.

Preliminary analyses were carried out to study demographic and genetic outcomes with varying parameters. These analyses enabled six variables of interest to be identified: the initial frequency of *avr* allele (f_{avr}), the migration rate (*mig*), the growth rate (r), the proportion of resistant hosts in the landscape (*propR*), the ploidy (*Ploidy*) and the life cycle (*Cycle*). Statistical analyses were performed on simulations with quantitative input parameters picked randomly from known distributions, resulting into a random simulation design (Table 1). The same simulation design was run four times, once for each combination of categorical input parameters (*Ploidy* and *Cycle*). To investigate further the impact of the input parameters on the simulation outcome in specific cases and to present the model results in a more didactic form, a regular simulation design was developed to complement the random design (Table 1).

This regular simulation design allowed us to present the results in a more conventional form. For both the random and the regular simulation design, simulations were run for each combination of parameters for 100 years (1100 generations) with 100 replicates. During this period, nearly all replicates reached equilibrium (fixation of one allele or extinction of the population).

A principal component analysis was performed on the data obtained with the random simulation design using R (R, 2018), on the following output variables: the frequency of extinction of

population (*freq_ext*), the frequency of fixation of the *Avr* allele in the population (*freq_fix_Avr*), the frequency of fixation of the *avr* allele (*freq_fix_avr*) and the generation of fixation of *avr* (*gen_fix_avr*). To study the influence of the six input parameters (*Ploidy*, *Cycle*, f_{avr}, *mig*, *r*, and *propR*) on the three main output variables selected (*freq_ext*, *freq_fix_avr*, and *gen_fix_avr*), generalized linear models (GLM) were performed on R. GLM on *freq_ext* and *freq_fix_avr* were performed with a Logistic link function, and the GLM on *gen_fix_avr* was performed with a Gamma link function.

To analyse further the temporal dynamics of *avr* allele frequency and population size, simulations were run recording population size and allelic states over time. Because these simulations were time- and memory-consuming, they were run on a restricted simulation design with only 24 combinations of parameters (Table 1). The generation of fixation of the *avr* allele was thus decomposed into two distinct output variables: the year of invasion of the R compartment and the time elapsed between the invasion and the *avr* allele fixation in the population. The influence of three parameters (*propR*, *Ploidy* and *Cycle*) on these two output variables was studied with 1000 replicates for each combination through 1100 generations. For these two output variables, GLM were performed with a Gamma link function.

For each general linear model developed, a dominance analysis was performed with the R package "dominanceanalysis" (Bustos Navarrete and Coutinho Soares, 2020) to compare the relative importance of the input parameters on the five output variables described. Estimated general dominance were calculated using bootstrap average values with the corresponding standard errors for each predictor with 100 bootstrap resamples, with McFadden's indices (McFadden, 1974).

Calculations of a growth rate threshold r_0 were carried out on Python for several parameter combinations. This value determines the growth rate below which the population goes extinct before the end of the simulation if there are no virulent individuals, therefore if the R compartment remains empty.

Table 1 – Input parameters and their range of variations for the three simulation designs. For both the random and regular simulation designs, 100 replicates were run for each combination to analyse the equilibrium reached by the population of pathogens after 1100 generations. For the restricted simulation design, 1000 replicates were run for each combination and the allele frequencies and population sizes were monitored through all 1100 generations simulated.

Variable	Description	Random design		Regular design	Restricted design
		Distribution	Interval		
f_{avr}	Initial frequency of the *avr* allele in the pathogen's population	Log-uniform	[log(0.0005), log(0.3)]	**6 levels**: 0.0005; 0.002; 0.01; 0.05; 0.15; 0.3	0.02
mig	Migration rate between R and S compartments	Log-uniform	[log(0.001), log(0.2)]	**3 levels**: 0.05; 0.1; 0.2	**2 levels**: 0.001; 0.05
r	Growth rate of the pathogen	Uniform	[1.1, 2]	**10 levels** : 1.1; 1.2; 1.3; 1.4; 1.5; 1.6; 1.7; 1.8; 1.9; 2.0	1.5
propR	Proportion of resistant hosts in the landscape	Uniform	]0, 1[	**9 levels**: 0.05; 0.1; 0.2; 0.35; 0.5; 0.65; 0.8; 0.9; 0.95	**3 levels**: 0.1; 0.5; 0.9
Ploidy	Ploidy of the pathogen	-	Haploid or Diploid	**2 levels**: Haploid; Diploid	**2 levels**: Haploid; Diploid
Cycle	Life cycle of the pathogen	-	Without or with host alternation	**2 levels**: Without host alternation; With host alternation	**2 levels**: Without host alternation; With host alternation

3. Results

3.1. Model behaviour.

Since the model leads to stochastic outputs, we first analysed model behaviour in order to identify sound output variables. We visualised both population size and allele frequency dynamics through generations. The trajectory of each simulation either lead to the extinction of the entire pathogen population or to the fixation of one allele, provided that simulations last long enough. An example of such model behaviour is illustrated in Annex B (Figure S2) with four replicates, assuming diploid pathogens with host alternation. In this example, population sizes display cyclical dynamics due to the annual migration event on the A compartment. Three out of five replicates lead to population extinctions, while in the two other replicates, some pathogen individuals succeed to invade the R compartment after the initial phase of population dynamics collapse, leading to the fixation of the *avr* allele. These two dynamics with the survival of the population following a genetic adaptation to harsh environment illustrates evolutionary rescue. Interestingly, all replicates succeed to invade the R compartment at some time, but - because of host alternation - the annual redistribution of individuals breaks the invasion dynamics of the R compartment. Therefore, invasion does not necessarily lead to *avr* fixation.

In the following, we will summarise simulation results with four output variables, computed over replicates: the frequency of extinction, *freq_ext*; the frequency of fixation of *Avr* or *avr* allele, *freq_fix_Avr* or *freq_fix_avr*, respectively; and the generation of fixation of *avr* allele, *gen_fix_avr*. The later output variable provides insights on the durability of resistance.

3.2. Sensitivity analyses.

To identify the most significant parameters on the different output variables, we conducted a PCA analysis, general linear models and dominance analyses.

The PCA analysis was performed on the four output variables, with the first and second axes accounting for 49.5% and 33.3% of the total variability respectively (Figure 2). The most influential parameters of interest on the output variables were the growth rate r, negatively correlated with the frequency of extinction of population *freq_ext*. The initial frequency of *avr* alleles in the population f_{avr} was positively correlated with the frequency of fixation of the *avr* allele *freq_fix_avr*. The migration rate *mig* and the proportion of resistant hosts in the landscape *propR* were negatively correlated with both the frequency of fixation of the *Avr* allele *freq_fix_Avr*, and the generation of fixation of the *avr* allele *gen_fix_avr*. The qualitative input parameters (*Ploidy* and *Cycle*) were studied by representing each of the combinations of parameters of the random simulation design colored according to its ploidy and life cycle (Figure 2.b). This PCA highlights a higher frequency of extinctions of population for diploids with host alternation. Moreover, simulations without host alternation lead to higher frequencies of fixation of *Avr*, and longer times to *avr* fixation. The ellipses also illustrate that haploid individuals with host alternation lead to less variable outcomes and to higher frequencies of fixation of *avr*.

Dominance analyses highlight the high impact of r on *freq_ext* (Figure 3.a), which is confirmed by the analysis of Sobol indices (Annex A Figure S1). For *freq_fix_avr* and *gen_fix_avr*, the influence of model parameters is more balanced with a lower contribution of the input parameters on *gen_fix_avr* (Figure 3.b, c). Overall, this analysis points out that *freq_ext* and *freq_fix_avr* are relatively well explained while *gen_fix_avr* is more stochastic.

3.3. Patterns of virulence fixation.

Three different and exclusive equilibria are observed: the extinction of the pathogen population, the fixation of the *avr* allele and the fixation of the *Avr* allele. The frequencies of these outputs among replicates are represented depending on f_{avr}, r, *propR* and *Cycle*, for haploids and diploids (Figure 4, Annex B Figure S5). For both ploidy levels, there is an increase in the frequency of fixation of the *avr* allele with the increase in f_{avr} and r. This representation also highlights the existence of a growth rate threshold r_0 above which there is fixation of either the *avr* allele or the *Avr* allele, and below which there is instead either fixation of the *avr* allele or extinction of the population. In other words, for a growth rate below r_0 the pathogen population only survives

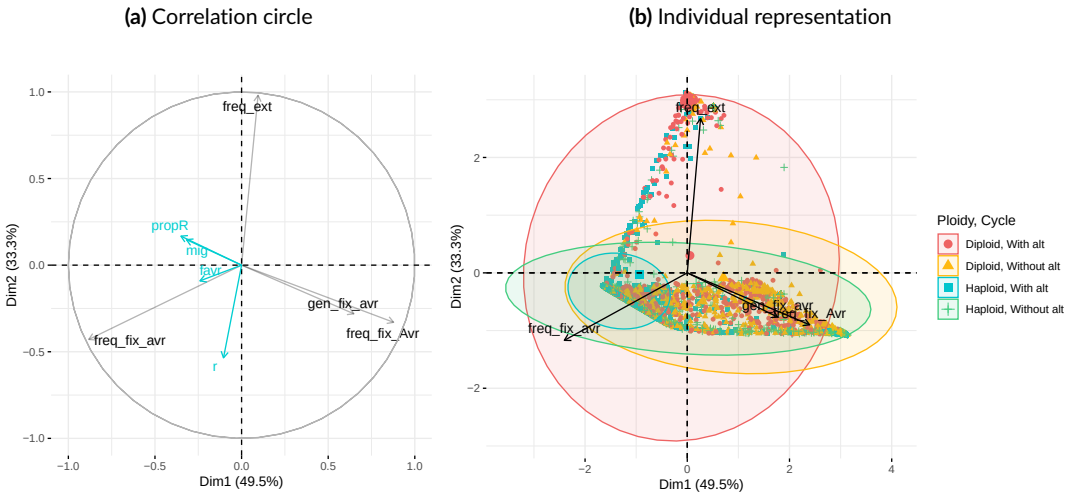


Figure 2 – Principal component analysis on four output variables: *freq_ext*, *freq_fix_Avr*, *freq_fix_avr*, and *gen_fix_avr*. Two results are displayed: (a) the correlation circle on four output variables. The quantitative parameters f_{avr}, *mig*, *r*, and *propR* are represented informatively in blue on the correlation circle and do not contribute to the variability; (b) the individual representation of simulations which represents the influence of the pathogen ploidy and life cycle. Each point represents a different combination of parameters with 100 replicates. Ellipses correspond to the 95% multivariate distribution.

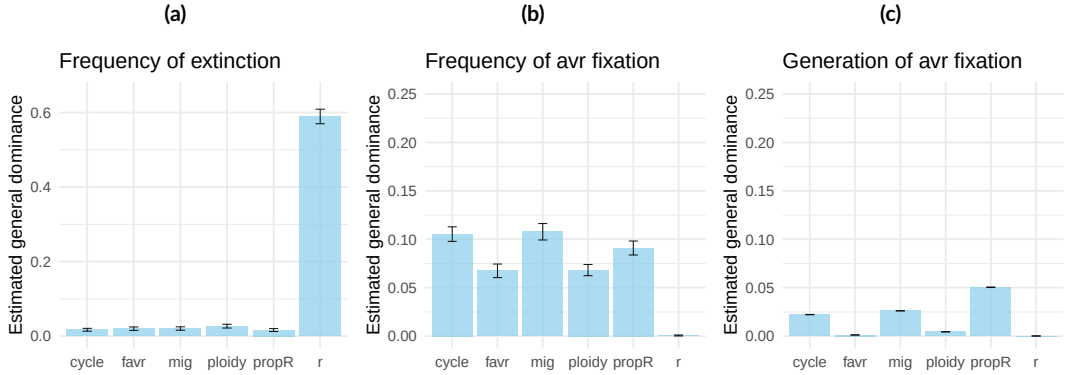


Figure 3 – Estimated general dominance of each predictor calculated from general linear models applied to three output variables of the random simulation design: the frequency of extinction of population, the frequency of fixation of the *avr* allele among replicates with surviving populations, and the generation of fixation of the *avr* allele. For each predictor the general dominance was estimated from bootstrap average values with the corresponding standard error for 100 bootstrap resamples.

when virulent individuals invade the R compartment, which corresponds to evolutionary rescue. Evolutionary rescue is particularly noticeable for the life cycle with host alternation because in this case, r_0 increases with *propR*.

Above r_0, we observe a gradient between the predominant fixation of the two alleles depending of f_{avr}, with slightly different patterns influenced by *propR*, *Cycle* and *r* for diploids (Figure 4). The influence of the life cycle on the fixation pattern is the most noticeable for low values of *propR* and *r*. The frequency of fixation of the *avr* allele appears maximal for intermediate values of *propR*.

To examine further the influence of *propR* on the probability of fixation of the *avr* allele, we plotted the evolution of the frequency of fixation of the *avr* allele among all replicates of the regular simulation design depending on *propR* and *r* for a fixed value of f_{avr} (Figure 5). For haploid individuals with host alternation, the frequency of fixation of the *avr* allele increases very slightly with *propR*. For haploids without host alternation, a plateau is observed for intermediate values of *propR*. For diploids, the distribution is shifted with a peak of maximal proportion of *avr* fixation for a slightly lower value of *propR*.

3.4. Variations in the speed of virulence spread.

To analyse in more details the dynamics of virulence spread, we examine two time points, reflecting two measures of resistance durability: the invasion of the R compartment and the resistance breakdown event. Invasion and resistance breakdown were defined as the first year when at least 1‰ and 1% of the resistant compartment were occupied by pathogens, respectively. Distributions of these two measures of resistance durability were plotted for three values of *propR*, with and without host alternation, only for the replicates for which we observed eventually a fixation of the *avr* allele. To broaden the picture, we monitored also the evolutionary trajectory of the *avr* allele from the invasion of the R compartment. The dynamics of invasion is mainly driven by the ploidy level and the dynamics of virulence spread is mainly driven by *propR* (Annex B Figure S3).

For haploid individuals resistance breakdown occur very rapidly: during the first or second year of simulation, regardless of the life cycle (figure not shown). There is a small delay in the time of the resistance breakdown with the increase in *propR*, especially without host alternation.

For diploid individuals, we observed longer periods before invasion and resistance breakdown and higher kurtosis. Assuming a strong migration rate, there are few differences between life cycles on the time of invasion. Both life cycles display an increase in kurtosis that goes hand in hand with the increase in *propR*. Without host alternation, distributions of the year of resistance breakdown and invasion time are similar, but with a one-year lag. Conversely, with host alternation, there are strong changes in the distributions that flatten out when considering the year of resistance breakdown (Figure 6). Assuming a low migration rate (i.e. for telluric pathogens), distributions of the year of invasion and resistance breakdown remain unchanged with host alternation, while these distributions flatten out considerably without host alternation (Annex B Figure S4). Note that in both migration regimes, with host alternation we observe a bimodal distribution of invasion year for $propR = 0.9$, with many invasion events in the first year of simulations. Early invasion events result from the initial redistribution of pathogen individuals following sexual reproduction on the alternate host: it is all the more likely that a virulent individual arrives on the R compartment the more predominant it is in the landscape.

In a last step, the evolution of the frequency of the *avr* allele in the population was studied along with the evolution of the population sizes through generations, from the invasion (Figure 7, see Annex B Figure S6 for haploids). For both life cycles, we observe an increase in the speed of fixation of the *avr* allele with the increase in *propR*. The median speed of fixation is higher with host alternation for haploids, and highly dependant of *propR* without host alternation for diploids (Annex B Figure S3). We also observe differences in stochasticity levels depending on the ploidy and the life cycle. For haploids, the dynamics of virulence fixation is almost deterministic. For diploids, the dynamics are more variable, with a highly stochastic behaviour for the life cycle with host alternation. Moreover, the results of GLM, the dominance analysis and the comparison of both figures show that independently of *propR* and of the life cycle, the increase in the proportion of *avr* allele is faster for haploid than for diploid individuals.

Focusing on diploid simulations, Figure 7.b highlights the existence of an evolutionary rescue effect, for the life cycle with host alternation and a high value of *propR*. The median population size decreases through time and almost reached 0 before the 20th generation following the invasion, when an increase in the proportion of *avr* alleles lead to an increase in the population size in the R compartment, followed by an increase in the population size of the S compartment, preventing the extinction of the population.

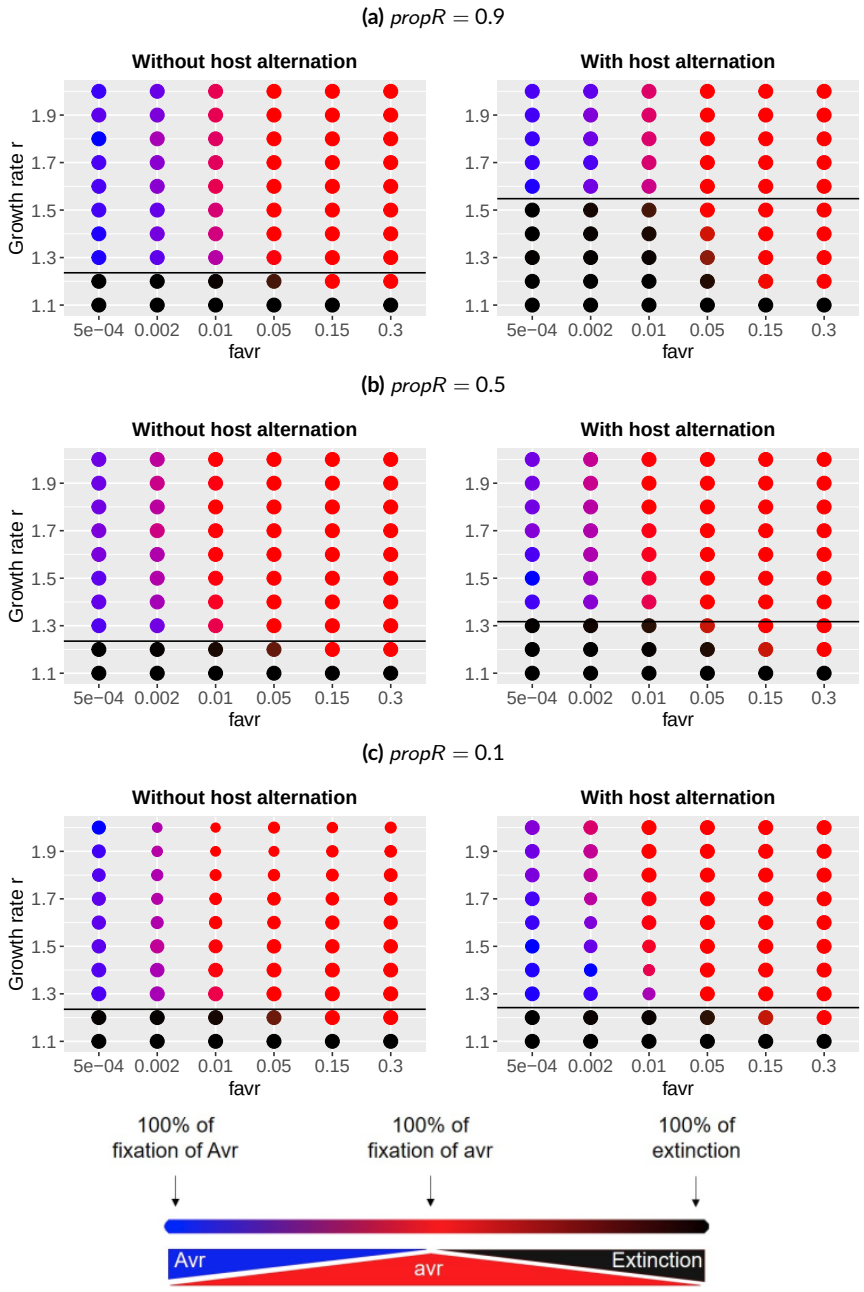


Figure 4 – For diploid pathogens, representation of the frequencies of population extinction or fixation of alleles *Avr* or *avr* for each combination of four parameters: f_{avr}, r, *propR* and *cycle*, with $mig = 0.05$. On each graph the black line corresponds to the calculated value of the growth rate threshold r_0 below which the population dies if it does not expand to the R compartment. The surface of each plotted result is proportional to the number of simulations, among the 100 replicates, for which an equilibrium was reached at the end of the 1100 generations simulated.

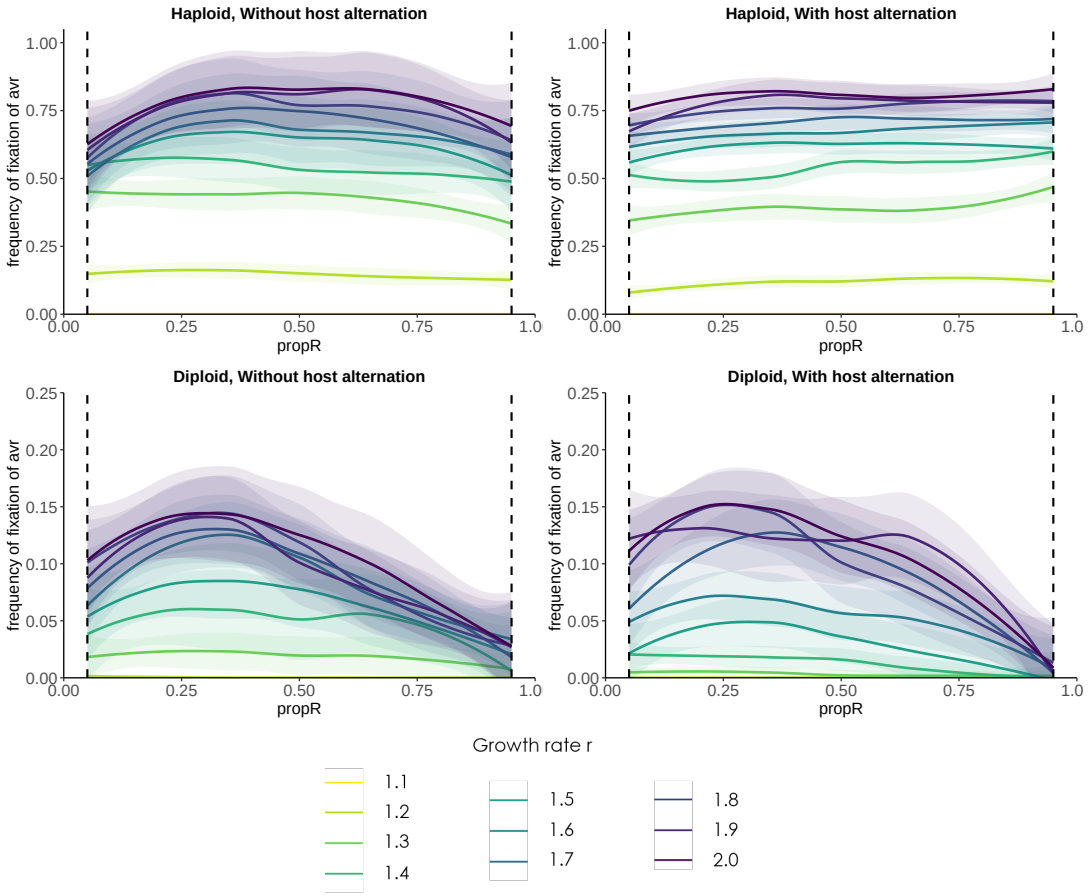


Figure 5 – Evolution of the frequency of fixation of the *avr* allele depending on *propR* for *r* varying between 1.1 and 2.0, with $f_{avr} = 0.0005$. Simulations were performed without and with host alternation, for haploid and diploid individuals, with 100 replicates for each combination of parameters. The plotted results correspond to the local regression on the frequency of fixation of the *avr* allele with the 95% confidence intervals associated with each regression. The vertical dotted lines correspond to the bounds of simulated values of *propR* for this regular experimental design.

Interestingly, the speed of invasion is mainly driven by the ploidy, while the speed of fixation of the *avr* allele from the invasion is mainly driven by the landscape composition *propR* (Annex B Figure S3).

4. Discussion

4.1. Deep impact of the ploidy on resistance durability.

Lof et al. (2017) demonstrated that the pre-existence of virulent alleles in the pathogen population could greatly diminish resistance durability. In the present study, we varied the initial frequency of *avr* alleles in the population but focused only on cases where this allele was already present at the beginning of the simulation, which corresponds to standing genetic variation (Alexander et al., 2014; Barrett et al., 2008). Our results illustrated a strong positive relationship between the initial frequency of virulent alleles and the probability of invasion and resistance breakdown. For haploid individuals, we found no stochasticity in the time of occurrence of the invasion, which occurred in the first year of the simulation for all replicates. Thus, for simulations with haploid pathogens, almost as soon as one virulent individual invaded the resistant compartment, it was selected and the resistance breakdown occurred. This result explains why a lot of

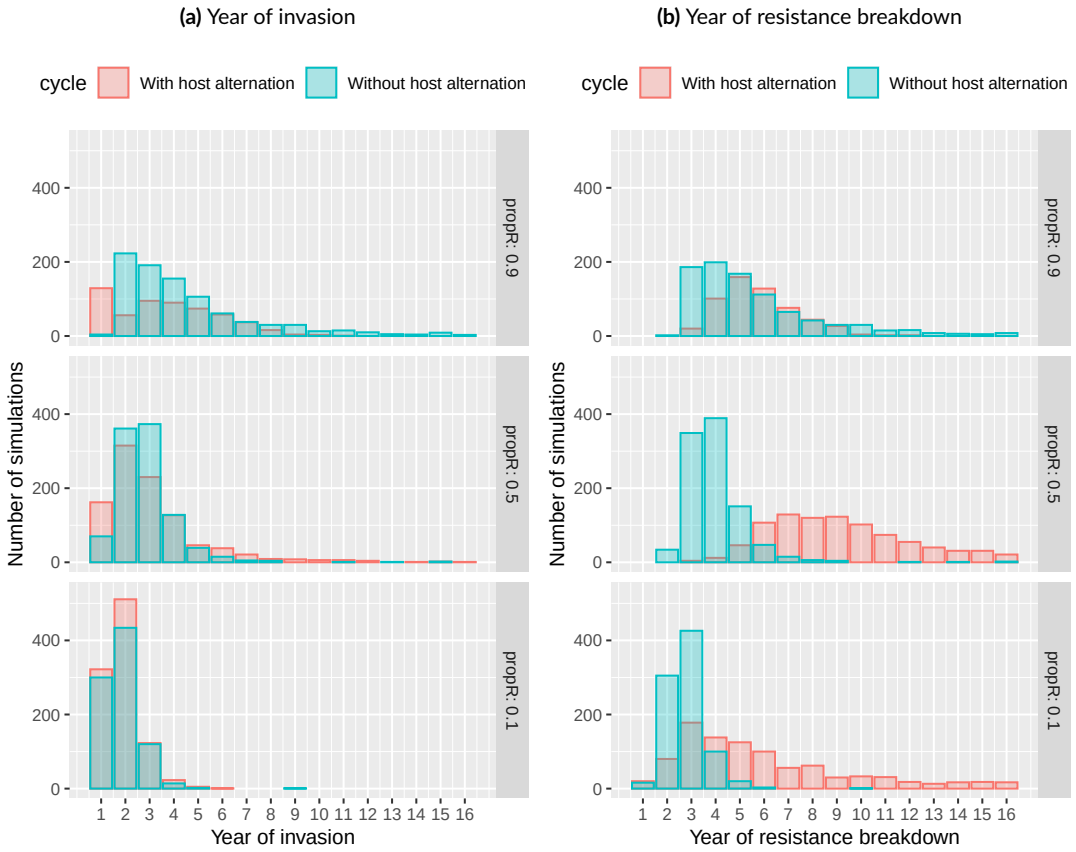


Figure 6 – Histograms of (a) the year of invasion and (b) the year of resistance breakdown depending on the life cycle of the diploid pathogen, for three values of *propR*. Simulations were performed with $f_{avr} = 0.02$, $mig = 0.05$, and $r = 1.5$. The plotted results were obtained from the restricted simulation design, and correspond to all simulations among the 1 000 replicates per combinations for which we observed a fixation of the *avr* allele in the population.

models on haploid individuals focus on the probability of apparition of the first virulent individual, in particular by mutation (Fabre et al., 2015; Papaïx et al., 2015). Our results for simulations with haploid pathogens also highlighted low stochasticity in the increase in the proportion of virulent alleles in the population after the invasion. The results obtained with haploids were consistent with previous studies on resistance durability (Pacilly et al., 2018), which permitted us to consider this model as the reference model, in order to study the influence of the diploid state on the system dynamics. Diploid individuals, however, display strongly different evolutionary trajectories. In particular, we observed higher stochasticities in the evolution of the virulent allele frequency, both before and after the resistance breakdown. This is mainly caused by the recessivity of the *avr* allele, according to the gene-for-gene model. Before the invasion, the *avr* allele is rare and mostly at the heterozygous state, hence leading to phenotypically avirulent individuals. Therefore, the *avr* allele is poorly selected, and variations in its frequency are mostly driven by genetic drift, which induces high stochasticity among replicates. This effect should be strengthened in species with small effective population sizes, such as in cyst nematodes (Gracianne et al., 2016; Montarry et al., 2019).

As a consequence, we observed lower frequencies of *avr* fixation and higher extinction rates for diploid individuals, independently of the life cycle and the host deployment strategy. Moreover, simulations with diploid pathogens resulted in lower speeds of fixation of the virulent allele, that is, higher resistance durability. Because of the heterozygous *Avr*/*avr* state, *avr* alleles are

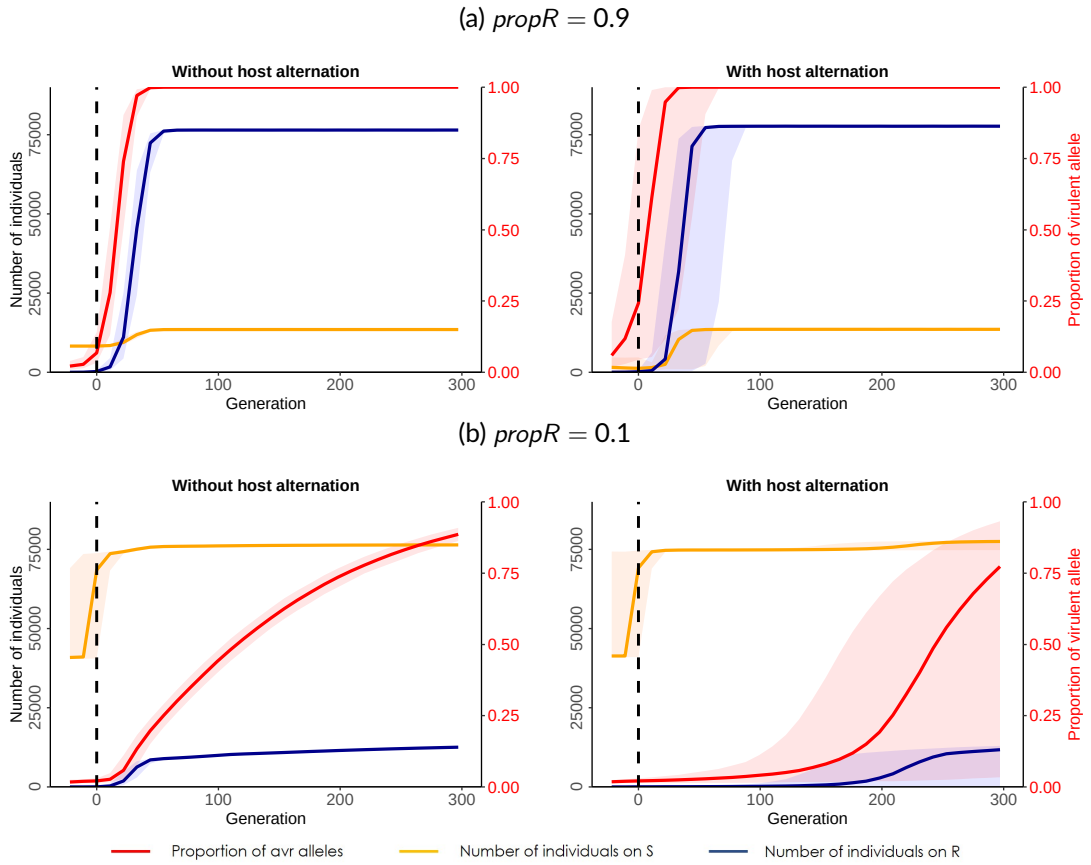


Figure 7 – For diploid pathogens, evolution of the population size in the R and S compartments (on the left scale) and of the frequency or *avr* alleles in the S compartment (on the right scale) through generations. Simulations were performed without and with host alternation, for two values of *propR*: (a) 0.1 and (b) 0.9, with $f_{avr} = 0.02$, $mig = 0.05$, and $r = 1.5$. For each simulation, generation 0 corresponds to the generation at which the invasion occurred. For each combination of parameters, simulations were performed on 1 000 replicates. The plotted results correspond to the median results (frequency of *avr* alleles or population size) for all simulations among the 1 000 replicates for which we observed a fixation of the *avr* allele in the population. Coloured intervals correspond to the 95% confidence intervals.

not necessarily selected and their presence in the population does not inevitably lead to an immediate resistance breakdown, as observed for haploid individuals. Thus, besides its impact on the stochasticity of the results, the vulnerability of the virulent allele at the heterozygous state is also responsible for an increase in resistance durability. The impact of the landscape composition on resistance durability also differs with the ploidy. Consistently with the work of Van den Bosch and Gilligan (2003) and Pietravalle et al. (2006), for haploid individuals the increase in *propR* leads to a strong increase in the speed of fixation of the *avr* allele, thus decreasing the resistance durability. In all cases except for haploids with host alternation, this result was accompanied in the present study by a maximum frequency of *avr* fixation for intermediate values of *propR*. This non-linear relationship is similar to the one highlighted for haploids by Pacilly et al. (2018), and is caused by two distinct mechanisms. At low proportions of resistant hosts in the landscape, the selective pressure on the *avr* allele is sufficiently low to limit the increase in the virulence in the pathogen population. At high values of *propR*, the opposition between selection and drift is magnified: on one hand the selective pressure is high and imposes a rapid pace of adaptation; on the other hand the small size of the S compartment increases genetic drift and

the risk of extinction of the *avr* allele, provided that the R compartment is not invaded. Hence, in most cases the virulent allele is lost if it does not spread quickly enough in the population: either $r > r_0$ which lead to a fixation of the avirulent *Avr* allele, either $r < r_0$ and the population goes extinct. Therefore, it would be possible to reduce the probability of invasion for diploid pathogens with either very low or very high proportions of resistant hosts in the landscape. However the increase in the proportion of resistant hosts is at the risk of weaker resistance durability: if the resistance breakdown occurs, it occurs more rapidly. For haploid pathogens with host alternation, we observed an almost constant and slightly increasing frequency of fixation of *avr* with the increase in *propR*. In this case and contrary to diploids with host alternation, the increase in the proportion of *avr* alleles on the resistant host is not counteracted by the allele reshuffling during the sexual reproduction event on the alternate host. For diploids, this allelic reshuffling causes a rise in the number of phenotypically avirulent *Avr*/*avr* heterozygous individuals, which will die if the redistribution following the sexual reproduction lead them on the resistant host. This can result in a drastic drop in the *avr* allele proportion while for haploids, the proportion of resistant hosts in the landscape does not increase the mortality rate of individuals carrying the virulent allele, because these haploid individuals are necessarily surviving on the resistant host.

4.2. Life cycles impose different selection regimes and lead to contrasted evolutionary trajectories.

The two different life cycles considered in this study - with or without host alternation - can be assimilated to hard and soft selection respectively (Christiansen, 1975; Wallace, 1975). Soft selection is expected to protect polymorphims, and hence promote local adaptation, while hard selection resembles an all or nothing game, that is to adapt to the encountered environment or to perish. Here host alternation can lead to faster evolution of allelic frequencies, with higher speeds of virulence fixation, especially for high values of *propR*. Without host alternation on the contrary, the evolution of virulence alleles are buffered, which result in more constrained dynamics. The increase in gene flow resulting from host alternation limits natural selection and local adaptation (Lenormand, 2002), especially because of the dispersal of non-adapted individuals on resistant hosts. The life cycle with host alternation is thus characterized by higher probabilities of population extinctions, and strong dependency of the growth rate threshold r_0 and the landscape composition *propR*. For diploids with host alternation, contrary to the local adaptation on each compartment without host alternation, the forced allelic reshuffling on the alternate host is responsible for the increase in the number of *Avr*/*avr* heterozygous individuals. Because the *avr* allele is recessive, a large proportion of these newly-produced individuals die from the redistribution on the resistant compartment following the sexual reproduction. Noticeably, the reduction in virulence fixation at high proportions of resistant host discussed above hence results from two mechanisms in diploids: impediment of local adaptation without host alternation or increase in selective pressure with host alternation.

For diploid individuals, we also observed contrasting patterns of variation in the evolution of the *avr* allele frequency before and after the invasion, depending on the life cycle. The time of invasion is more stochastic without host alternation, while the speed of increase in the *avr* allele frequency from the invasion is far more stochastic with host alternation. The first step relies essentially on the probability of encounter between a virulent individual and a resistant host. Without host alternation the encounter is restrained to the probability that a virulent individual migrates from the susceptible to the resistant host during the vegetative season. The host alternation reinforces gene-flow, with the annual migration event that redistributes pathogen individuals among all host plants, thereby favoring the encounter. Interestingly, in the case of host alternation, early infections of resistant host (invasion) does not readily translate into population establishment on that host (disease outbreak leading to resistance breakdown). At the end of the vegetative season and initial invasion, the sexual reproduction on alternate host reshuffles allele frequencies, and thus breaks virulent (homozygous) individuals into mostly avirulent (heterozygous) individuals. These up and down selection phases amplify the effect of genetic drift and lead to nearly chaotic evolutionary trajectories, resulting in a resistance durability all the more difficult to predict. Without host alternation, virulent individuals mate with each others

and the homozygous state is sheltered, which results in a strict time lag between initial invasion and population outbreak. Overall our model is in accordance with the framework proposed by McDonald and Linde (2002) which highlights the importance of gene flow as an impediment to resistance durability. Our analysis completes this framework, taking into account the variation in life cycles.

The life cycle also plays a role in the frequency of observation of evolutionary rescue effects. Carolan et al. (2017) highlighted the impact of the growth rate of the pathogen on resistance durability, by presenting the limitation of the growth rate as a mean to increase resistance durability. In accordance with this study, we displayed a growth rate threshold r_0 below which the pathogen population goes extinct if it does not invade the resistant compartment. Hence, for a growth rate below r_0, the genetic adaptation of the pathogen population is the only way for the population to survive, which is a classical example of evolutionary rescue. In the current study, r_0, and thus the observation range of evolutionary rescue, is highly dependent on the life cycle. Without host alternation, the redistribution of individuals between compartments and the mortality rate is limited, which leads to a quite low r_0, independently of the proportion of resistant hosts in the landscape. With host alternation, however, the redistribution event occurring each year from the alternate host to the S and R compartments leads to a strong dependence of r_0 on the landscape composition, with an increase in the observation range of evolutionary rescue with the proportion of resistant hosts.

5. Conclusion

Short-term epidemiological control is predicted to be optimal for a landscape composed of a high proportion of resistant hosts in a low degree of spatial aggregation (Holt and Chancellor, 1999; Papaïx et al., 2014b, 2018; Skelsey et al., 2010). However other authors also highlighted that optimal resistance durability could be obtained by reducing the proportion of resistant hosts (Fabre et al., 2012; Papaïx et al., 2018; Pietravalle et al., 2006; Pink and Puddephat, 1999), thus minimising the selection pressure on the pathogen population. In the current study, the minimisation of the probability of fixation of the virulent allele for a diploid pathogen population was obtained either at very low or very high proportions of resistant hosts in the landscape. In cases where the population does not go extinct and the virulent allele increases in proportion, however, the proportion of resistant hosts in the landscape strongly impacts the speed of increase, and thus the resistance durability. Consistently with Van den Bosch and Gilligan (2003), we displayed that for a diploid pathogen population with standing genetic variation, the increase in the proportion of resistant hosts decreases resistance durability. In particular, with host alternation both the invasion and the fixation of the virulent allele in the population can occur very quickly. However, in such a case, the evolutionary trajectory of the virulent allele is all the more stochastic and durability is thus difficult to predict. Without host alternation (i.e. for the majority of pathogen species) early detection and population control measures would increase resistance durability. However such prophylactic measures are made all the more difficult by the strong unpredictability of the invasion date. For the few species with host alternation, a massive intervention could durably control a population of pathogens, such as the eradication of the alternate host species *Berberis vulgaris* for wheat stem rust control (Peterson, 2018). Overall, the high stochasticity of evolutionary trajectory impedes accurate forecasts of resistance durability for diploid organisms.

6. Perspectives

In the current study, we considered a single qualitative resistance gene. The combination of several resistance genes is often studied and deployed to increase resistance durability (Djian-Caporalino et al., 2014; Djidjou-Demasse et al., 2017; Mundt, 2014; Rimbaud et al., 2018b). These combinations of resistances can occur at the plant scale with multiple resistance genes (pyramiding) inside one host genotype, or at the landscape scale with multiple resistance deployments in time or space (Mundt, 2002; Van den Bosch and Gilligan, 2003). Some resistant cultivars progressively introduced in the landscape are composed of different combinations of qualitative resistance genes, resulting in an evolving selective pressure through time (Goyeau

and Lannou, 2011). Building on our results, we can extrapolate on the impact of the ploidy and the life cycle on resistance durability for these different strategies of deployment. Hence, rotating cultures with different resistances would amount to force gene flow, favoring the encounter of pathogen individuals with new hosts without an actual migration. This would be comparable to the hard selection regime observed with host alternation, and we expect similar results. With the pyramiding of several resistance genes in same host, meanwhile, we would expect a higher short term efficiency than with a single resistance, but with a higher risk of rapid resistance breakdown by multi-virulent individuals due to the inducing of a strong selective pressure. This may especially be true for pathogens with host alternation because of the increased probability of mating between pathogens with different virulent profiles when they meet on the alternate host.

7. Data accessibility

Python simulation code and results, as well as R script for result analyses are available on Zenodo repository: (DOI: 10.5281/zenodo.4892587).

8. Acknowledgements

We warmly thank Josselin Montarry, Lydia Bousset, Jean-Paul Soularue, Frédéric Grognard, Bénédicte Fabre, Clémentine Louet, Pascal Frey, and Cyril Dutech for constructive comments on previous versions of the manuscript. We thank Frédéric Fabre for insightful discussions on sensibility analyses. We thank AgroParisTech for support. We also thank Hirohisa Kishino, Loup Rimbaud and one anonymous reviewer for detailed comments and suggestions which helped improve the manuscript. A previous version of this article has been peer-reviewed and recommended by Peer Community In Evolutionary Biology (DOI: 10.24072/pci.evolbiol.100131).

9. Funding

This work was supported by grants from the French National Research Agency (ANR-18-CE32-0001, Clonix2D project; ANR-11-LABX-0002-01, Cluster of Excellence ARBRE). Méline Saubin was supported by a PhD fellowship from INRAE and the French National Research Agency (ANR-18-CE32-0001, Clonix2D project).

10. Conflict of interest disclosure

The authors of this article declare that they have no financial conflict of interest with the content of this article.

Annex A: Sobol indices

Sensitivity analyses were performed with the calculation of Sobol indices (first-order, second-order and total-order) with the R package "sensitivity" (Iooss et al., 2021). Sobol indices were calculated to study the importance of each of the six parameters of interest on the output variable *freq_ext* only (figure S1). These calculations were based on the results issued from the random simulation design.

For the four remaining outputs (*freq_fix*, *gen_fix*, the year of occurrence of the invasion, and the time elapsed between the invasion and the fixation of the *avr* allele), only the simulations not leading to extinction were retained for the sensitivity analyses. Thus, the input combinations of parameters retained depended strongly on the results of the output variable *freq_ext*, the independence hypothesis of the input parameters were then not verified and Sobol indices could not be calculated for these four remaining output variables. Further analyses would be necessary to disentangle the effect of each parameter of interest on these remaining output variables.

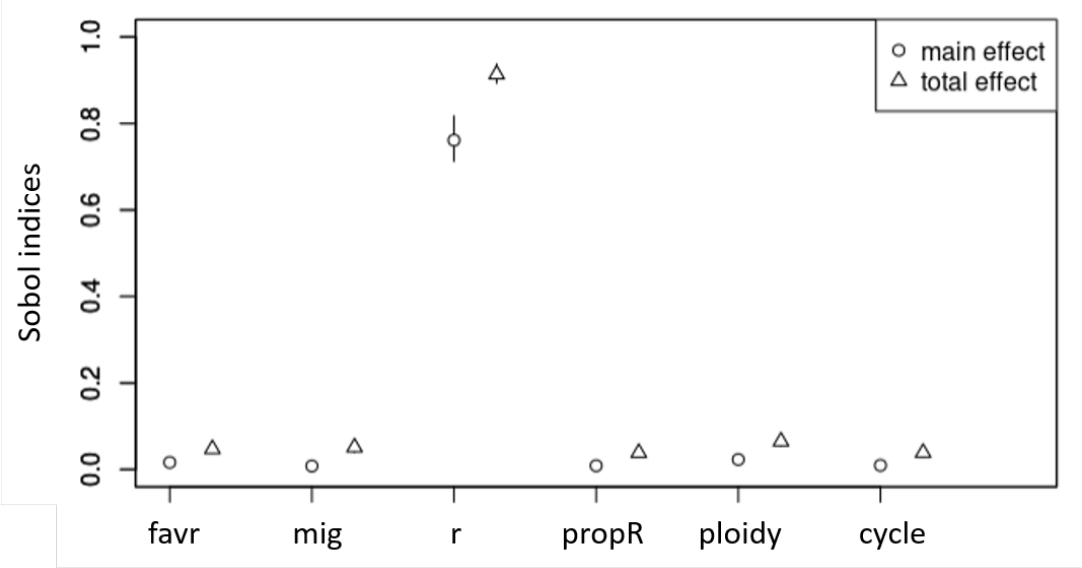


Figure S1 – Sobol indices to evaluate the influence of six variables of interest on the frequency of extinctions among simulations. Main effect correspond to first-order Sobol indices, and total effect correspond to total-order Sobol indices.

Annex B: Supplementary figures

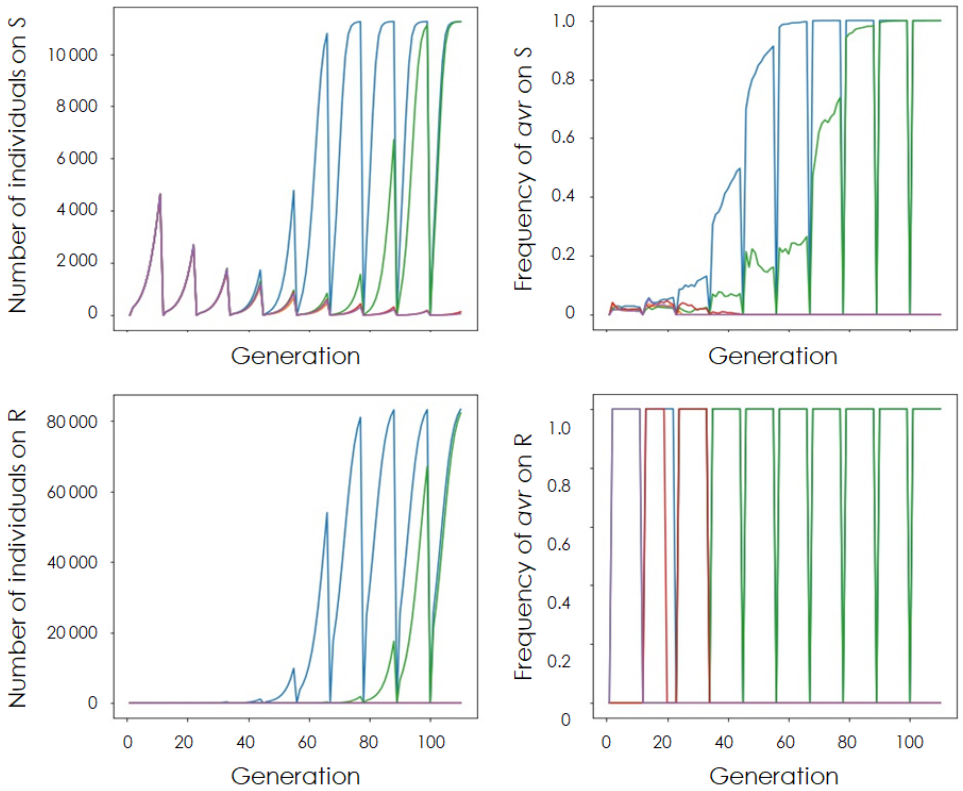


Figure S2 – Example of the simulated populations demographic (on the left) and virulent allele frequency (on the right) dynamics through time on S and R compartments. The model was run for four replicates with diploid individuals and host alternation, $propR = 0.9$, $r = 1.5$, $f_{avr} = 0.025$, and $mig = 0.05$. Each color represents a distinct replicate.

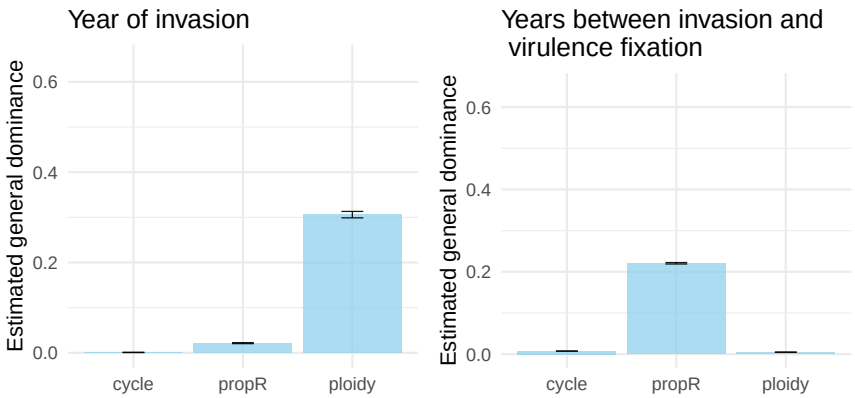


Figure S3 – Estimated general dominance of each predictor calculated from general linear models applied to two output variables of the restricted simulation design: the year of invasion and the time elapsed between the invasion and the fixation of the *avr* allele. For each predictor the general dominance was estimated from bootstrap average values with the corresponding standard error for 100 bootstrap resamples.

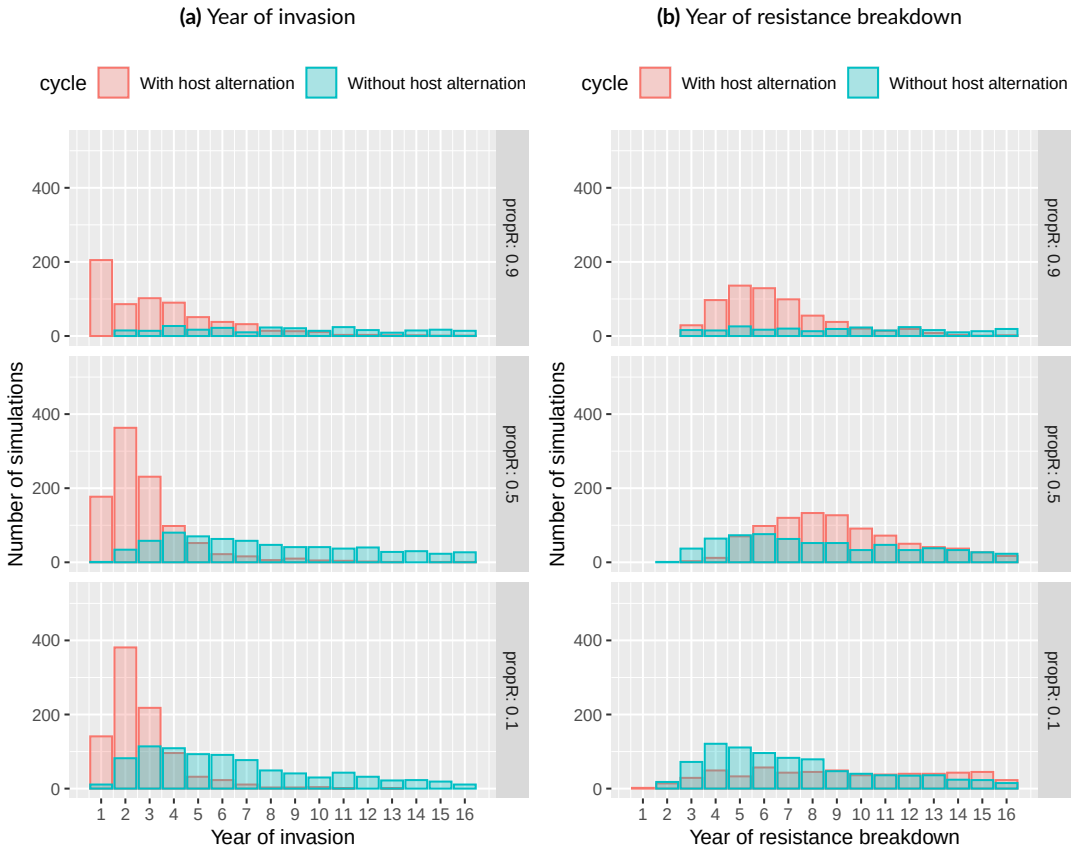


Figure S4 – Histograms of (a) the year of invasion and (b) the year of resistance breakdown depending on the life cycle of the diploid pathogen, for three values of *propR*. Simulations were performed with $f_{avr} = 0.02$, $mig = 0.001$, and $r = 1.5$. The plotted results were obtained from the restricted simulation design, and correspond to all simulations among the 1 000 replicates per combinations for which at least 80% of the R compartment is occupied at the end of the simulation.

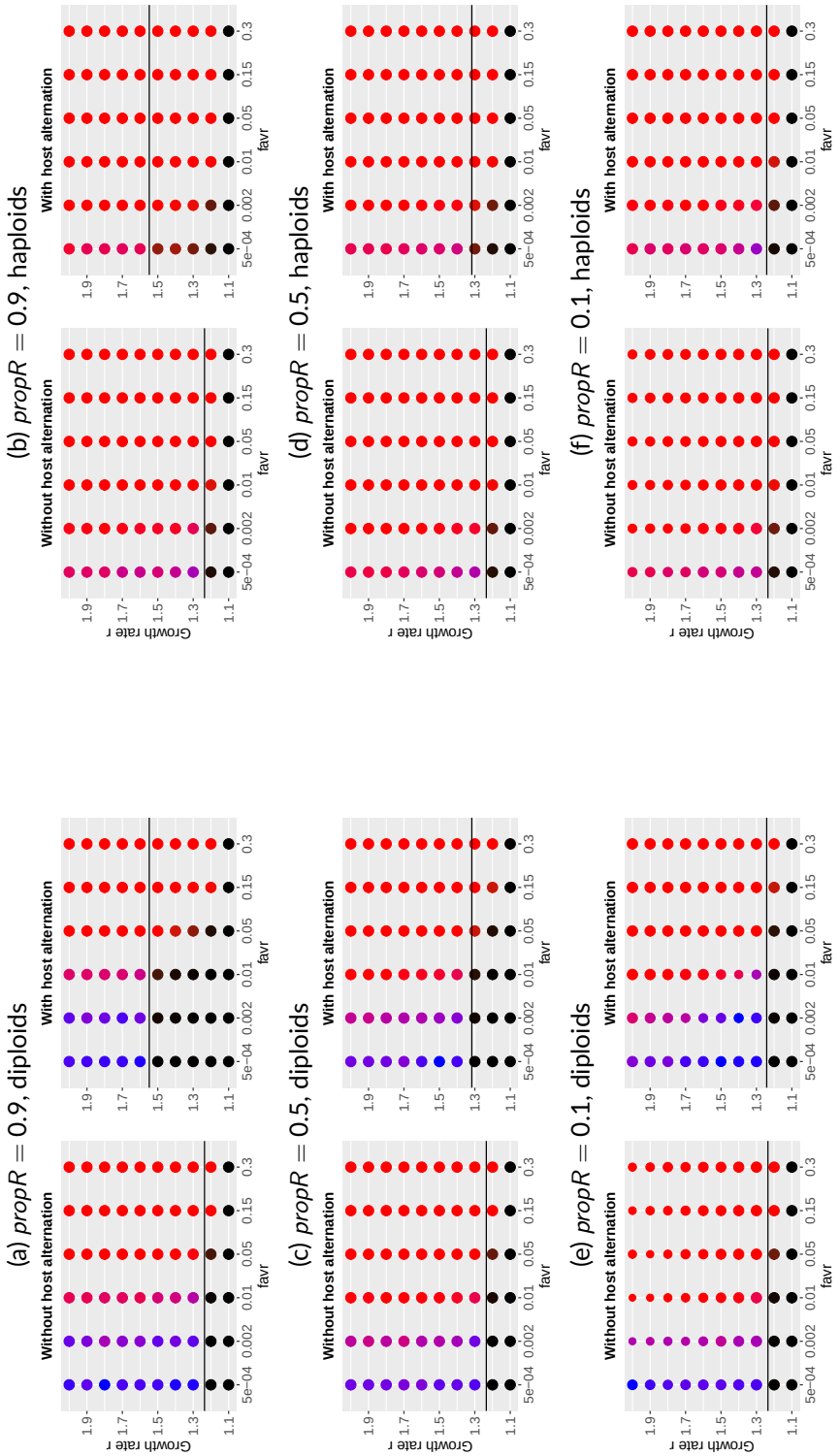


Figure S5 – Representation of the frequencies of population extinction or fixation of alleles *Avr* or *avr* for each combination of five parameters: f_{avr}, *r*, *propR*, *ploidy* and *cycle*, with $mig = 0.05$. On each graph the black line corresponds to the calculated value of the growth rate threshold r_0 below which the population dies if it does not expand to the R compartment. The surface of each plotted result is proportional to the number of simulations, among the 100 replicates, for which an equilibrium was reached at the end of the 1100 generations simulated. In colored dots, red corresponds to the fixation of the *avr* allele, blue to the fixation of the *Avr* allele, and black to the extinction of the population. Dot color shades indicate simulation results among replicates. The left part ((a), (c), and (e)) corresponds to Figure 4, presented in the main document.

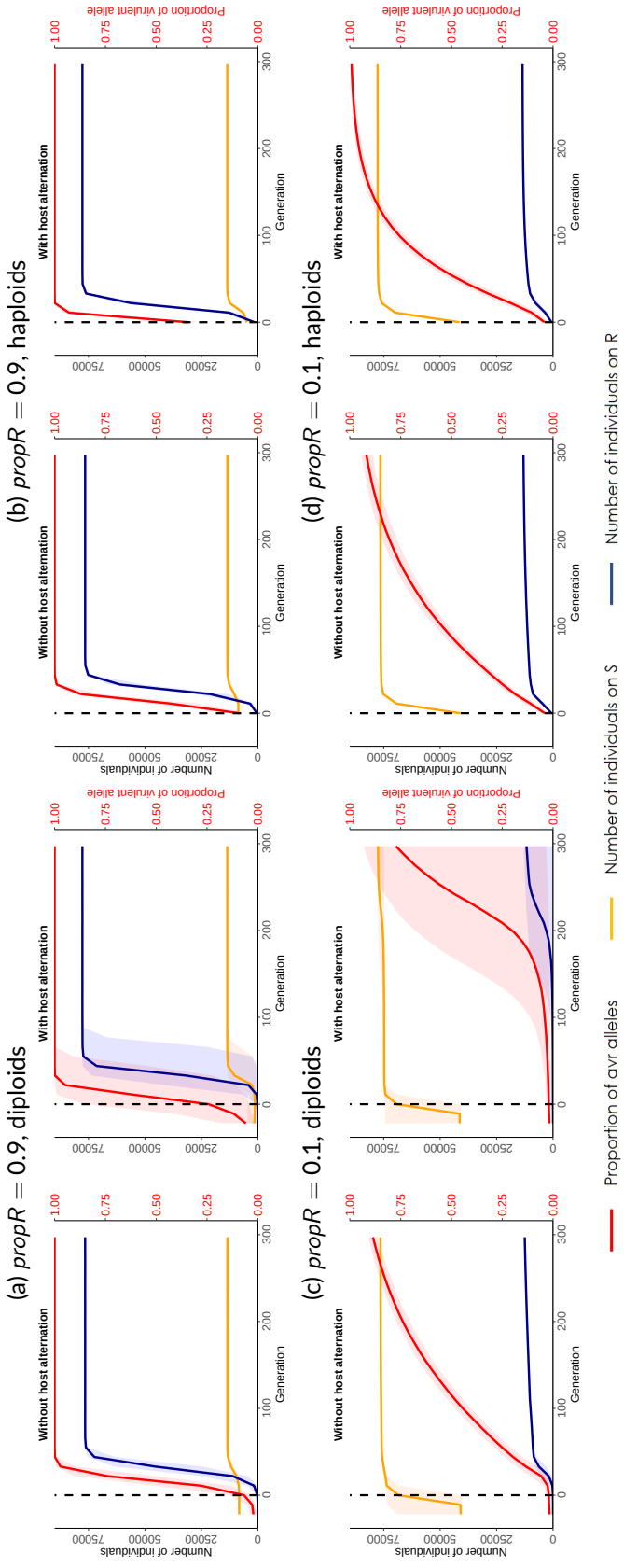


Figure S6 – Evolution of the population size in the R and S compartments (on the left scale) and of the frequency or *avr* alleles in the S compartment (on the right scale) through generations. Simulations were performed without and with host alternation, for (a, c) diploid and (b, d) haploid pathogens, for two values of *propR*: (a, b) 0.9 and (c, d) 0.1, with $f_{avr} = 0.02$, $mig = 0.05$, and $r = 1.5$. For each simulation, generation 0 corresponds to the generation at which the invasion occurred. For each combination of parameters, simulations were performed on 1 000 replicates. The plotted results correspond to the median results (frequency of *avr* alleles or population size) for all simulations among the 1 000 replicates for which we observed a fixation of the *avr* allele in the population. Coloured intervals correspond to the 95% confidence intervals. The left part ((a) and (c)) corresponds to Figure 7, presented in the main document.

References

Agrios G (2005). *Plant pathology*. Elsevier. ISBN: 9780120445653.

Alexander HK et al. (2014). *Evolutionary rescue: Linking theory for conservation and medicine*. *Evolutionary Applications* **7**, 1161–1179. ISSN: 17524571. https://doi.org/10.1111/eva.12221.

Barrett LG et al. (2008). *Life history determines genetic structure and evolutionary potential of host-parasite interactions*. *Trends in Ecology and Evolution* **23**, 678–685. ISSN: 01695347. https://doi.org/10.1016/j.tree.2008.06.017.

Bazin É et al. (2014). *The effect of mating system on invasiveness: Some genetic load may be advantageous when invading new environments*. *Biological Invasions* **16**, 875–886. ISSN: 13873547. https://doi.org/10.1007/s10530-013-0544-6.

Berngruber TW et al. (2013). *Evolution of virulence in emerging epidemics*. *PLoS Pathogens* **9**. ISSN: 15537366. https://doi.org/10.1371/journal.ppat.1003209.

Bolker BM et al. (2010). *Transient virulence of emerging pathogens*. *Journal of the Royal Society Interface* **7**. ISSN: 17425662. https://doi.org/10.1098/rsif.2009.0384.

Bousset L, Chèvre AM (2013). *Stable epidemic control in crops based on evolutionary principles: Adjusting the metapopulation concept to agro-ecosystems*. *Agriculture, Ecosystems and Environment* **165**, 118–129. ISSN: 01678809. https://doi.org/10.1016/j.agee.2012.12.005. URL: http://dx.doi.org/10.1016/j.agee.2012.12.005.

Bousset L et al. (2018). *Spatio-temporal connectivity and host resistance influence evolutionary and epidemiological dynamics of the canola pathogen Leptosphaeria maculans*. *Evolutionary Applications* **11**, 1354–1370. ISSN: 17524571. https://doi.org/10.1111/eva.12630.

Brown JKM (2015). *Durable resistance of crops to disease: a darwinian perspective*. *Annual Review of Phytopathology* **53**, 513–539. ISSN: 0066-4286. https://doi.org/10.1146/annurev-phyto-102313-045914.

Brown JKM, Tellier A (2011). *Plant-parasite coevolution: bridging the gap between genetics and ecology*. *Annual Review of Phytopathology* **49**, 345–367. ISSN: 0066-4286. https://doi.org/10.1146/annurev-phyto-072910-095301.

Burdon JJ, Thrall PH (2008). *Pathogen evolution across the agro-ecological interface: implications for disease management*. *Evolutionary Applications* **1**, 57–65. ISSN: 1752-4571. https://doi.org/10.1111/j.1752-4571.2007.00005.x.

Burdon JJ et al. (2014). *Guiding deployment of resistance in cereals using evolutionary principles*. *Evolutionary Applications* **7**, 609–624. ISSN: 17524571. https://doi.org/10.1111/eva.12175.

Burdon JJ et al. (2016). *Addressing the challenges of pathogen evolution on the world's arable crops*. *Phytopathology* **106**, 1117–1127. ISSN: 0031949X. https://doi.org/10.1094/PHYTO-01-16-0036-FI.

Bustos Navarrete C, Coutinho Soares F (2020). *dominanceanalysis: Dominance Analysis*. R package version 2.0.0. URL: https://CRAN.R-project.org/package=dominanceanalysis.

Carolan K et al. (2017). *Extending the durability of cultivar resistance by limiting epidemic growth rates*. *Proceedings of the Royal Society B: Biological Sciences* **284**. ISSN: 14712954. https://doi.org/10.1098/rspb.2017.0828.

Chen W et al. (2014). *Wheat stripe (yellow) rust caused by Puccinia striiformis f. sp. tritici*. *Molecular Plant Pathology* **15**, 433–446. ISSN: 13643703. https://doi.org/10.1111/mpp.12116.

Christiansen FB (1975). *Hard and soft selection in a subdivided sopulation*. *The American Naturalist* **109**, 11–16.

Condeso TE, Meentemeyer RK (2007). *Effects of landscape heterogeneity on the emerging forest disease sudden oak death*. *Journal of Ecology* **95**, 364–375. ISSN: 00220477. https://doi.org/10.1111/j.1365-2745.2006.01206.x.

Day T, Gandon S (2007). *Applying population-genetic models in theoretical evolutionary epidemiology*. *Ecology Letters* **10**, 876–888. ISSN: 1461023X. https://doi.org/10.1111/j.1461-0248.2007.01091.x.

Day T, Proulx SR (2004). *A general theory for the evolutionary dynamics of virulence. The American naturalist* **163**. ISSN: 15375323. https://doi.org/10.1086/382548.

Djian-Caporalino C et al. (2014). *Pyramiding, alternating or mixing: Comparative performances of deployment strategies of nematode resistance genes to promote plant resistance efficiency and durability. BMC Plant Biology* **14**, 1–13. ISSN: 14712229. https://doi.org/10.1186/1471-2229-14-53.

Djidjou-Demasse R et al. (2017). *Mosaics often outperform pyramids: Insights from a model comparing strategies for the deployment of plant resistance genes against viruses in agricultural landscapes. New Phytologist*, 239–253. ISSN: 14698137. https://doi.org/10.1111/nph.14701.

Fabre F et al. (2012). *Durable strategies to deploy plant resistance in agricultural landscapes. New Phytologist* **193**, 1064–1075. ISSN: 0028646X. https://doi.org/10.1111/j.1469-8137.2011.04019.x.

Fabre F et al. (2015). *Epidemiological and evolutionary management of plant resistance: Optimizing the deployment of cultivar mixtures in time and space in agricultural landscapes. Evolutionary Applications* **8**, 919–932. ISSN: 17524571. https://doi.org/10.1111/eva.12304.

Flor HH (1971). *Current status of the gene-for-gene concept. Annual Review of Phytopathology* **9**, 275–296. ISSN: 0066-4286. https://doi.org/10.1146/annurev.py.09.090171.001423.

Garrett KA et al. (2009). *Intraspecific functional diversity in hosts and its effect on disease risk across a climatic gradient. Ecological Applications* **19**, 1868–1883. ISSN: 10510761. https://doi.org/10.1890/08-0942.1.

Gilligan CA, Bosch F (2008). *Epidemiological models for invasion and persistence of pathogens. Annual Review of Phytopathology* **46**, 385–418. ISSN: 0066-4286. https://doi.org/10.1146/annurev.phyto.45.062806.094357.

Goyeau H, Lannou C (2011). *Specific resistance to leaf rust expressed at the seedling stage in cultivars grown in France from 1983 to 2007. Euphytica* **178**, 45–62. ISSN: 00142336. https://doi.org/10.1007/s10681-010-0261-5.

Gracianne C et al. (2016). *Temporal sampling helps unravel the genetic structure of naturally occurring populations of a phytoparasitic nematode. 2. Separating the relative effects of gene flow and genetic drift. Evolutionary Applications* **9**, 1005–1016. ISSN: 17524571. https://doi.org/10.1111/eva.12401.

Griette Q et al. (2015). *Virulence evolution at the front line of spreading epidemics. Evolution* **69**, 2810–2819. ISSN: 15585646. https://doi.org/10.1111/evo.12781.

Haas SE et al. (2011). *Forest species diversity reduces disease risk in a generalist plant pathogen invasion. Ecology Letters* **14**, 1108–1116. ISSN: 1461023X. https://doi.org/10.1111/j.1461-0248.2011.01679.x.

Holt J, Chancellor TCB (1999). *Modelling the spatio-temporal deployment of resistant varieties to reduce the incidence of rice tungro disease in a dynamic cropping system. Plant Pathology* **48**, 453–461. ISSN: 00320862. https://doi.org/10.1046/j.1365-3059.1999.00360.x.

Huang YJ et al. (2006). *Fitness cost associated with loss of the AvrLm4 avirulence function in Leptosphaeria maculans (phoma stem canker of oilseed rape). European Journal of Plant Pathology* **114**, 77–89. ISSN: 09291873. https://doi.org/10.1007/s10658-005-2643-4.

Huang YJ et al. (2010). *Fitness cost of virulence differs between the AvrLm1 and AvrLm4 loci in Leptosphaeria maculans (phoma stem canker of oilseed rape). European Journal of Plant Pathology* **126**, 279–291. ISSN: 09291873. https://doi.org/10.1007/s10658-009-9539-7.

Iooss B et al. (2021). *sensitivity: Global Sensitivity Analysis of Model Outputs*. R package version 1.24.0. URL: https://CRAN.R-project.org/package=sensitivity.

Johnson R (1984). *A critical analysis of durable resistance. Annual Review of Phytopathology* **22**, 309–330. ISSN: 0042-3106. https://doi.org/10.1007/bf00047316.

Laine AL, Barrès B (2013). *Epidemiological and evolutionary consequences of life-history trade-offs in pathogens. Plant Pathology* **62**, 96–105. ISSN: 00320862. https://doi.org/10.1111/ppa.12129.

Langlois JP et al. (2001). *Landscape structure influences continental distribution of hantavirus in deer mice. Landscape Ecology* **16**, 255–266. ISSN: 09212973. https://doi.org/10.1023/A:1011148316537.

Leach JE et al. (2001). *Pathogen fitness penalty as a predictor of durability of disease resistance genes*. *Annual Review of Ecology, Evolution, and Systematics* **39**, 187–224.

Lenormand T (2002). *Gene flow and the limits to natural selection*. *Trends in Ecology and Evolution* **17**, 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7. arXiv: S0169-5347(01)02432-6.

Lewontin RC (1958). *A general method for investigating the equilibrium of gene frequency in a population*. *Genetics* **43**, 419–434.

Lo Iacono G et al. (2012). *The evolution of plant pathogens in response to host resistance: Factors affecting the gain from deployment of qualitative and quantitative resistance*. *Journal of Theoretical Biology* **304**, 152–163. ISSN: 00225193. https://doi.org/10.1016/j.jtbi.2012.03.033. URL: http://dx.doi.org/10.1016/j.jtbi.2012.03.033.

Lo Presti L et al. (2015). *Fungal effectors and plant susceptibility*. *Annual Review of Plant Biology* **66**, 513–545. ISSN: 15452123. https://doi.org/10.1146/annurev-arplant-043014-114623.

Lof ME et al. (2017). *Achieving durable resistance against plant diseases: Scenario analyses with a national-scale spatially explicit model for a wind-dispersed plant pathogen*. *Phytopathology* **107**, 580–589. ISSN: 0031949X. https://doi.org/10.1094/PHYTO-05-16-0207-R.

Lof ME, Werf W (2017). *Modelling the effect of gene deployment strategies on durability of plant resistance under selection*. *Crop Protection* **97**, 10–17. ISSN: 02612194. https://doi.org/10.1016/j.cropro.2016.11.031. URL: http://dx.doi.org/10.1016/j.cropro.2016.11.031.

Lorrain C et al. (2019). *Advances in understanding obligate biotrophy in rust fungi*. *New Phytologist* **222**, 1190–1206. ISSN: 14698137. https://doi.org/10.1111/nph.15641.

Martin G et al. (2013). *The probability of evolutionary rescue: towards a quantitative comparison between theory and evolution experiments*. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**. ISSN: 14712970. https://doi.org/10.1098/rstb.2012.0088.

McDonald BA (2004). *Population genetics of plant pathogens*. *The Plant Health Instructor*. ISSN: 1935-9411. https://doi.org/10.1094/PHI-A-2004-0524-01. URL: https://www.apsnet.org/edcenter/advanced/topics/PopGenetics/Pages/default.aspx.

McDonald BA, Linde C (2002). *Pathogen population genetics , evolutionary potential, and durable resistance*. *Annual Review of Phytopathology* **40**, 349–379. ISSN: 0066-4286. https://doi.org/10.1146/annurev.phyto.40.120501.101443.

McFadden D (1974). *Conditional logit analysis of qualitative choice behavior*. In: *Frontiers in econometrics*, pp. 104–142.

Montarry J et al. (2012). *Virus adaptation to quantitative plant resistance: erosion or breakdown?* *Journal of Evolutionary Biology* **25**, 2242–2252. ISSN: 1010061X. https://doi.org/10.1111/j.1420-9101.2012.02600.x.

Montarry J et al. (2010). *Fitness costs associated with unnecessary virulence factors and life history traits: evolutionary insights from the potato late blight pathogen Phytophthora infestans*. *BMC Evolutionary Biology* **10**. https://doi.org/10.1186/1471-2148-10-283.

Montarry J et al. (2019). *Exploring the causes of small effective population sizes in cyst nematodes using artificial Globodera pallida populations*. *Proceedings of the Royal Society B: Biological Sciences* **286**. ISSN: 14712954. https://doi.org/10.1098/rspb.2018.2359.

Moran NA (1992). *The evolution of aphid life cycles*. *Annual review of entomology* **37**, 321–348. ISSN: 00664170. https://doi.org/10.1146/annurev.ento.37.1.321.

Mundt CC (2002). *Use of multiline cultivars and cultivar mixtures for disease management*. *Annual Review of Phytopathology* **40**, 381–410. ISSN: 0066-4286. https://doi.org/10.1146/annurev.phyto.40.011402.113723.

Mundt CC (2014). *Durable resistance: A key to sustainable management of pathogens and pests*. *Infection, Genetics and Evolution* **27**, 1567–1348. ISSN: 15677257. https://doi.org/10.1016/j.meegid.2014.01.011. URL: http://dx.doi.org/10.1016/j.meegid.2014.01.011.

Nilusmas S et al. (2020). *Multiseasonal modelling of plant-nematode interactions reveals efficient plant resistance deployment strategies*. *Evolutionary Applications*, 1–16. ISSN: 1752-4571. https://doi.org/10.1111/eva.12989.

Oleksyk TK et al. (2010). *Genome-wide scans for footprints of natural selection*. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 185–205. https://doi.org/10.1098/rstb.2009.0219. arXiv: rstb.2009.0219 [10.1098].

Ostfeld RS, Keesing F (2012). *Effects of host diversity on infectious disease*. *Annual Review of Ecology, Evolution, and Systematics* **43**, 157–182. ISSN: 1543-592X. https://doi.org/10.1146/annurev-ecolsys-102710-145022.

Pacilly FC et al. (2018). *Simulating crop-disease interactions in agricultural landscapes to analyse the effectiveness of host resistance in disease control: The case of potato late blight*. *Ecological Modelling* **378**, 1–12. ISSN: 03043800. https://doi.org/10.1016/j.ecolmodel.2018.03.010. URL: https://doi.org/10.1016/j.ecolmodel.2018.03.010.

Papaïx J et al. (2013). *Dynamics of adaptation in spatially heterogeneous metapopulations*. *PLoS ONE* **8**. ISSN: 19326203. https://doi.org/10.1371/journal.pone.0054697.

Papaïx J et al. (2014a). *Can epidemic control be achieved by altering landscape connectivity in agricultural systems? Ecological Modelling* **284**, 35–47. ISSN: 03043800. https://doi.org/10.1016/j.ecolmodel.2014.04.014. URL: http://dx.doi.org/10.1016/j.ecolmodel.2014.04.014.

Papaïx J et al. (2014b). *Pathogen population dynamics in agricultural landscapes: The Ddal modelling framework*. *Infection, Genetics and Evolution* **27**, 509–520. ISSN: 15677257. https://doi.org/10.1016/j.meegid.2014.01.022.

Papaïx J et al. (2015). *Crop pathogen emergence and evolution in agro-ecological landscapes*. *Evolutionary Applications* **8**, 385–402. ISSN: 17524571. https://doi.org/10.1111/eva.12251.

Papaïx J et al. (2018). *Differential impact of landscape-scale strategies for crop cultivar deployment on disease dynamics, resistance durability and long-term evolutionary control*. *Evolutionary Applications* **11**, 705–717. ISSN: 17524571. https://doi.org/10.1111/eva.12570.

Park RF (2008). *Breeding cereals for rust resistance in Australia*. *Plant Pathology* **57**, 591–602. ISSN: 00320862. https://doi.org/10.1111/j.1365-3059.2008.01836.x.

Peng B, Kimmel M (2005). *simuPOP: A forward-time population genetics simulation environment*. *Bioinformatics* **21**, 3686–3687. ISSN: 13674803. https://doi.org/10.1093/bioinformatics/bti584.

Perkins TA et al. (2013). *Evolution of dispersal and life history interact to drive accelerating spread of an invasive species*. *Ecology Letters* **16**. ISSN: 14610248. https://doi.org/10.1111/ele.12136.

Person C et al. (1962). *The gene-for-gene concept*. *Nature* **194**, 561–562.

Persoons A (2015). *Les contournements de résistance par Melampsora larici-populina, l'agent de la rouille du peuplier : impact démographique et déterminisme génétique*. PhD thesis. Université de Lorraine.

Persoons A et al. (2017). *The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust*. *Molecular Ecology*. https://doi.org/10.1111/mec.13980. arXiv: 0608246v3 [arXiv:physics].

Peterson PD (2018). *The barberry eradication program in Minnesota for stem rust control: a case study*. *Annual Review of Phytopathology* **56**, 203–223. ISSN: 15452107. https://doi.org/10.1146/annurev-phyto-080417-050133.

Pietravalle S et al. (2006). *Durability of resistance and cost of virulence*. *European Journal of Plant Pathology* **114**, 107–116. ISSN: 09291873. https://doi.org/10.1007/s10658-005-3479-7.

Pilet-Nayel ML et al. (2017). *Quantitative resistance to plant pathogens in pyramiding strategies for durable crop protection*. *Frontiers in Plant Science* **8**, 1–9. ISSN: 1664462X. https://doi.org/10.3389/fpls.2017.01838.

Pink D, Puddephat I (1999). *Deployment of disease resistance genes by plant transformation - a 'mix and match' approach*. *Trends in Plant Science* **4**, 71–75. ISSN: 13601385. https://doi.org/10.1016/S1360-1385(98)01372-7.

Plantegenest M et al. (2007). *Landscape epidemiology of plant diseases*. *Journal of the Royal Society Interface* **4**, 963–972. ISSN: 17425689. https://doi.org/10.1098/rsif.2007.1114.

R (2018). *R Core Team, R: A language and environment for statistical computing*.

Real LA, Biek R (2007). *Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes*. *Journal of the Royal Society Interface* **4**, 935–948. ISSN: 17425689. https://doi.org/10.1098/rsif.2007.1041.

Rimbaud L et al. (2018a). *Assessing the durability and efficiency of landscape-based strategies to deploy plant resistance to pathogens*. *PLoS Computational Biology* **14**, 1–33. ISSN: 15537358. https://doi.org/10.1371/journal.pcbi.1006067.

Rimbaud L et al. (2018b). *Mosaics, mixtures, rotations or pyramiding: What is the optimal strategy to deploy major gene resistance? Evolutionary Applications* **11**, 1791–1810. ISSN: 17524571. https://doi.org/10.1111/eva.12681.

Rimbaud L et al. (2021). *Models of plant resistance deployment*. *Annual Review of Phytopathology* **59**.

Rossum G (1995). *Python tutorial, Technical Report CS-R9526*. *CWI*.

Rousseau E (2017). *Effet de la dérive génétique et de la sélection sur la durabilité de la résistance des plantes aux virus*. PhD thesis.

Rouxel T, Balesdent MH (2017). *Life, death and rebirth of avirulence effectors in a fungal pathogen of Brassica crops, Leptosphaeria maculans*. *New Phytologist* **214**, 526–532. ISSN: 14698137. https://doi.org/10.1111/nph.14411.

Sacristán S, García-Arenal F (2008). *The evolution of virulence and pathogenicity in plant pathogen populations*. *Molecular Plant Pathology* **9**, 369–384. ISSN: 14646722. https://doi.org/10.1111/j.1364-3703.2007.00460.x.

Sapoukhina N et al. (Dec. 2009). *Spatial deployment of gene-for-gene resistance governs evolution and spread of pathogen populations*. *Theoretical Ecology* **2**, 229–238. ISSN: 1874-1738. https://doi.org/10.1007/s12080-009-0045-5. URL: http://link.springer.com/10.1007/s12080-009-0045-5.

Skelsey P et al. (2010). *Invasion of Phytophthora infestans at the landscape level: How do spatial scale and weather modulate the consequences of spatial heterogeneity in host resistance? Phytopathology* **100**, 1146–1161. ISSN: 0031949X. https://doi.org/10.1094/PHYTO-06-09-0148.

Soularue JP et al. (2017). *Short rotations in forest plantations accelerate virulence evolution in root-rot pathogenic fungi*. *Forests* **8**. ISSN: 19994907. https://doi.org/10.3390/f8060205.

Stukenbrock EH, McDonald BA (2009a). *Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions*. *Molecular Plant-Microbe Interactions* **22**, 371–380. ISSN: 08940282. https://doi.org/10.1094/MPMI-22-4-0371.

Stukenbrock EH, McDonald BA (2009b). *Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions*. *Molecular Plant-Microbe Interactions* **22**, 371–380. ISSN: 08940282. https://doi.org/10.1094/MPMI-22-4-0371.

Tellier A, Brown JK (2007). *Stability of genetic polymorphism in host-parasite interactions*. *Proceedings of the Royal Society B: Biological Sciences* **274**, 809–817. ISSN: 14712970. https://doi.org/10.1098/rspb.2006.0281.

Tellier A, Brown JK (2009). *The influence of perenniality and seed banks on polymorphism in plant-parasite interactions*. *American Naturalist* **174**, 769–779. ISSN: 00030147. https://doi.org/10.1086/646603.

Terauchi R, Yoshida K (2010). *Towards population genomics of effector-effector target interactions*. *New Phytologist* **187**, 929–939. ISSN: 0028646X. https://doi.org/10.1111/j.1469-8137.2010.03408.x.

Thrall PH, Burdon JJ (2003). *Evolution of virulence in a plant host-pathogen metapopulation*. *Science* **299**, 1735–1737. ISSN: 00368075. https://doi.org/10.1126/science.1080070.

Thrall PH et al. (2016). *Epidemiological and evolutionary outcomes in gene-for-gene and matching allele models*. *Frontiers in Plant Science* **6**. ISSN: 1664462X. https://doi.org/10.3389/fpls.2015.01084.

Van den Bosch F, Gilligan CA (2003). *Measures of durability of resistance*. *Phytopathology* **93**, 616–625. ISSN: 0031-949X. https://doi.org/10.1094/phyto.2003.93.5.616.

Wallace B (1975). *Hard and soft selection revisited*. *Evolution* **29**, 465–473.

Xhaard C (2011). *Influence des processus démographiques sur la structure et les caractéristiques génétiques des champignons phytopathogènes, cas de l'agent de la rouille du peuplier Melampsora larici-populina*. PhD thesis.

Zhan J et al. (2014). *Achieving sustainable plant disease management through evolutionary principles*. *Trends in Plant Science* **19**, 570–575. ISSN: 13601385. https://doi.org/10.1016/j.tplants.2014.04.010. URL: http://dx.doi.org/10.1016/j.tplants.2014.04.010.

Zhan J et al. (2015). *Playing on a aathogen's weakness: using evolution to guide sustainable plant disease control strategies*. *Annual Review of Phytopathology* **53**, 19–43. ISSN: 0066-4286. https://doi.org/10.1146/annurev-phyto-080614-120040.

2.2 Analyse déterministe par résolution analytique

2.2.1 Enjeux de la résolution déterministe

Pour comprendre plus précisément les équilibres des fréquences alléliques au cours d'un contournement de résistance, je complète l'analyse précédente par une résolution analytique déterministe du système dans le cas où l'agent pathogène est diploïde. Pour cela, le formalisme mathématique d'un modèle semi-discret correspond parfaitement au cycle de vie de l'agent pathogène étudié (MAILLERET ET LEMESLE, 2009; HAMELIN *et al.*, 2011). Le cycle de vie de Mlp présente chaque année de nombreux cycles de multiplication clonale pendant une saison de végétation, ce qui correspondrait à la partie continue du modèle, suivie d'un unique évènement de reproduction sexuée qui correspondrait à la partie discrète du modèle. Modéliser ces deux parties continue et discrète peut cependant se révéler compliqué. Dans cette thèse, la partie discrète du modèle a posé des problèmes de résolution analytique. Je présente donc ici les résultats de l'analyse de la partie continue du modèle uniquement. Cette analyse permet d'étudier les équilibres des fréquences alléliques au cours d'une saison épidémique clonale.

Ce modèle me permet, dans les cas les plus simples, de déterminer les différents équilibres, les fréquences de génotypes associées sur chaque compartiment d'hôtes, ainsi que les conditions d'existence et de stabilité de ces équilibres. Je compare ensuite ces résultats avec les différents équilibres retrouvés dans la partie précédente multi-annuelle et stochastique. Cela me permet de discuter l'impact de la stochasticité et du brassage génétique annuel lors de la reproduction sexuée sur l'évolution des agents pathogènes étudiés.

Ce travail a fait l'objet d'une collaboration avec Frédéric Grognard et Suzanne Touzeau et leur doctorante Clotilde Djuikem, tous trois travaillant à l'Inria Sophia-Antipolis.

2.2.2 Description du modèle

Ce modèle analytique représente l'évolution de la quantité d'agents pathogènes de chaque génotype G évoluant à l'instant t sur chaque compartiment d'hôtes H. H s'écrit S pour l'hôte sensible, et R pour l'hôte résistant. Au locus d'avirulence, A et a représentent respectivement l'allèle avirulent et l'allèle virulent. Chaque agent pathogène diploïde peut donc présenter un génotype G parmi ces trois combinaisons : AA, Aa, ou aa.

Ce modèle est composé de deux phases temporelles distinctes : une phase temporelle continue qui correspond à la phase de reproduction clonale sur l'espèce d'hôte télien et une phase temporelle discrète qui correspond à la reproduction sexuée sur l'hôte alternant. Nous cherchons à déterminer les différents équilibres lors de la reproduction clonale, leurs conditions d'existence et de stabilité. Nous décrirons donc la phase continue du modèle.

Au cours de la phase temporelle continue sur l'hôte télien, deux quantités sont obser-

vées au cours du temps t : la quantité de mycélium (en tant qu'unité d'infection) d'agents pathogènes de génotype G sur l'hôte H (M_G^H, équation (2.1)), et la quantité de spores de génotype G sur l'hôte H (U_G^H, équation (2.2)). La quantité de mycélium augmente de manière logistique (traduisant le processus d'infection par de nouvelles spores) et diminue par mortalité. La quantité de spores augmente par la sporulation du mycélium et diminue par la conversion de spores en mycélium lors de l'infection :

$$\frac{dM_G^H}{dt} = \beta_G^H \, f^H(U_G^H) \, (1 - \frac{M^H}{\kappa^H}) - \mu \, M_G^H \tag{2.1}$$

$$\frac{dU_G^H}{dt} = \gamma_G^H \, M_G^H - \omega_G^H \, U_G^H \tag{2.2}$$

avec β_G^H le succès d'implantation des individus de génotype G sur l'hôte H, κ^H la capacité de charge du compartiment H, M^H la quantité totale de mycélium des individus sur l'hôte H, μ le taux de mortalité du mycélium, γ_G^H le taux de sporulation des individus de génotype G sur l'hôte H, ω_G^H le taux de conversion des spores de génotype G sur l'hôte H en mycélium (Tableau 2.1).
Pour chaque compartiment d'hôtes, cette fonction d'infection par les spores $f^H(U_G^H)$ s'exprime :

$$f^S(U_G^S) = \alpha \, \omega_G^S \, U_G^S + (1 - \alpha) \, \omega_G^R \, U_G^R$$

$$f^R(U_G^R) = \alpha \, \omega_G^R \, U_G^R + (1 - \alpha) \, \omega_G^S \, U_G^S$$

avec α la proportion de spores émises qui restent sur leur compartiment d'origine.

TABLEAU 2.1 – Paramètres utilisés dans le modèle ainsi que leur domaine de définition. H représente le compartiment d'hôtes : S pour le compartiment d'hôtes sensibles ou R pour le compartiment d'hôtes résistants. G représente le génotype : AA, Aa, ou aa avec A l'allèle avirulent et a l'allèle virulent.

Paramètre	Description	Domaine
M_G^H	Quantité totale de mycélium des individus de génotype G sur l'hôte H	$[0, \infty]$
M^H	Quantité totale de mycélium des individus sur l'hôte H	$[0, \infty]$
U_G^H	Nombre total de spores des individus de génotype G sur l'hôte H	$[0, \infty]$
β_G^H	Succès d'implantation des individus de génotype G sur l'hôte H	$[0, 1]$
γ_G^H	Taux de sporulation des individus de génotype G sur l'hôte H	$[0, 1]$
ω_G^H	Taux de transformation des spores de génotype G sur l'hôte H en mycélium	$[0, 1]$
κ^H	Capacité de charge du compartiment H	$]0, \infty]$
α	Proportion des spores émises qui restent sur leur compartiment d'origine	$[0, 1]$
μ	Taux de mortalité du mycélium ($1/\mu$ = temps de sporulation)	$]0, 1[$

La dynamique d'évolution des spores est considérée comme rapide par rapport à la dynamique d'évolution de la quantité de mycélium. Nous pouvons donc nous placer au quasi-équilibre correspondant :

$$\frac{dU_G^H}{dt} = 0$$

Ce qui implique d'après (2.2) :

$$U_G^H = \frac{\gamma_G^H \, M_G^H}{\omega_G^H}$$

En remplaçant dans (2.1), l'évolution des quantités de mycélium au cours du temps sur les compartiments d'hôtes S et R s'expriment :

$$\frac{dM_G^S}{dt} = \beta_G^S \left(\alpha \, \gamma_G^S \, M_G^S + (1-\alpha) \, \gamma_G^R \, M_G^R\right) \left(1 - \frac{M^S}{\kappa^S}\right) - \mu \, M_G^S \tag{2.3}$$

$$\frac{dM_G^R}{dt} = \beta_G^R \left(\alpha \, \gamma_G^R \, M_G^R + (1-\alpha) \, \gamma_G^S \, M_G^S\right) \left(1 - \frac{M^R}{\kappa^R}\right) - \mu \, M_G^R \tag{2.4}$$

Nous obtenons donc un système d'équations issu de l'emboîtement du système lent-rapide décrivant les dynamiques de sporulation et d'augmentation du nombre d'unités mycéliennes d'infection au cours de la saison épidémique.

Nous souhaitons étudier l'évolution de la quantité d'agents pathogènes (mycélium) de chaque génotype sur les différents compartiments S et R. Considérant que l'allèle de virulence a est récessif, les individus de génotype AA et Aa sont phénotypiquement avirulent et ne peuvent se développer que sur le compartiment S, tandis que les individus de génotype aa peuvent se développer sur S et R. Nous pouvons donc décliner les équations (2.3) et (2.4) représentant l'évolution de la quantité de mycélium pour chaque génotype :

$$\begin{cases} \dfrac{dM_{AA}^S}{dt} = \alpha \, \beta_{AA}^S \, \gamma_{AA}^S \, M_{AA}^S \left(1 - \dfrac{M^S}{\kappa^S}\right) - \mu \, M_{AA}^S & (f_1) \\ \dfrac{dM_{Aa}^S}{dt} = \beta_{Aa}^S \left(\alpha \, \gamma_{Aa}^S \, M_{Aa}^S\right) \left(1 - \dfrac{M^S}{\kappa^S}\right) - \mu \, M_{Aa}^S & (f_2) \\ \dfrac{dM_{aa}^S}{dt} = \beta_{aa}^S \left(\alpha \, \gamma_{aa}^S \, M_{aa}^S + (1-\alpha) \, \gamma_{aa}^R \, M_{aa}^R\right) \left(1 - \dfrac{M^S}{\kappa^S}\right) - \mu \, M_{aa}^S & (f_3) \\ \dfrac{dM_{aa}^R}{dt} = \beta_{aa}^R \left(\alpha \, \gamma_{aa}^R \, M_{aa}^R + (1-\alpha) \, \gamma_{aa}^S \, M_{aa}^S\right) \left(1 - \dfrac{M^R}{\kappa^R}\right) - \mu \, M_{aa}^R & (f_4) \end{cases} \tag{2.5}$$

Dans ce modèle, nous faisons l'hypothèse que l'allèle avirulent A confère un avantage sélectif sur S, *via* un meilleur succès d'implantation des spores et un meilleur taux de

sporulation :

$\beta^S_{AA} > \beta^S_{Aa} > \beta^S_{aa}$

$\gamma^S_{AA} > \gamma^S_{Aa} > \gamma^S_{aa}$

Remarque : aucune condition sur β^R_{aa} et γ^R_{aa} n'est posée en dehors de leur domaine de définition.

À partir des quatre équations du système (2.5), nous nous intéressons à 16 conditions d'équilibre distinctes.

1. La disparition de l'ensemble des génotypes
2. La disparition des génotypes Aa
3. La disparition des génotypes AA
4. La disparition des génotypes Aa et aa sur S et R
5. La disparition des génotypes AA, Aa et aa sur R
6. La disparition des génotypes AA, Aa et aa sur S
7. La disparition des génotypes AA et Aa
8. La disparition des génotypes AA et aa sur S et R
9. La disparition des génotypes AA et aa sur S
10. La disparition des génotypes AA et aa sur R
11. La disparition des génotypes Aa et aa sur S
12. La disparition des génotypes Aa et aa sur R
13. La disparition des génotypes aa sur S et R
14. La disparition des génotypes aa sur S
15. La disparition des génotypes aa sur R
16. La survie de l'ensemble des génotypes

Pour ces conditions, nous souhaitons déterminer la quantité de mycélium aux équilibres pour chaque génotype sur chaque compartiment (points d'équilibres), ainsi que les conditions d'existence et de stabilité de ces équilibres.

2.2.3 Les différents équilibres et leur stabilité

La détermination des conditions de stabilité des équilibres nécessite le calcul des dérivées partielles du système, afin de construire les différentes matrices Jacobiennes. Les dérivées partielles obtenues à partir du système (2.5) sont les suivantes :

$$\begin{cases} \dfrac{\partial f_1}{\partial M^S_{AA}} = \alpha\, \beta^S_{AA}\, \gamma^S_{AA}\, (1 - \dfrac{M^S}{\kappa^S}) - \dfrac{1}{\kappa^S}\, \alpha\, \beta^S_{AA}\, \gamma^S_{AA}\, M^S_{AA} - \mu \\ \dfrac{\partial f_1}{\partial M^S_{Aa}} = -\dfrac{1}{\kappa^S}\, \alpha\, \beta^S_{AA}\, \gamma^S_{AA}\, M^S_{AA} \\ \dfrac{\partial f_1}{\partial M^S_{aa}} = -\dfrac{1}{\kappa^S}\, \alpha\, \beta^S_{AA}\, \gamma^S_{AA}\, M^S_{AA} \\ \dfrac{\partial f_1}{\partial M^R_{aa}} = 0 \end{cases}$$

$$\begin{cases} \dfrac{\partial f_2}{\partial M^S_{AA}} = -\dfrac{1}{\kappa^S}\, \alpha\, \beta^S_{Aa}\, \gamma^S_{Aa}\, M^S_{Aa} \\ \dfrac{\partial f_2}{\partial M^S_{Aa}} = \alpha\, \beta^S_{Aa}\, \gamma^S_{Aa}\, (1 - \dfrac{M^S}{\kappa^S}) - \dfrac{1}{\kappa^S}\, \alpha\, \beta^S_{Aa}\, \gamma^S_{Aa}\, M^S_{Aa} - \mu \\ \dfrac{\partial f_2}{\partial M^S_{aa}} = -\dfrac{1}{\kappa^S}\, \alpha\, \beta^S_{Aa}\, \gamma^S_{Aa}\, M^S_{Aa} \\ \dfrac{\partial f_2}{\partial M^R_{aa}} = 0 \end{cases}$$

$$\begin{cases} \dfrac{\partial f_3}{\partial M^S_{AA}} = -\dfrac{1}{\kappa^S}\, \beta^S_{aa}\, (\alpha\, \gamma^S_{aa}\, M^S_{aa} + (1-\alpha)\, \gamma^R_{aa}\, M^R_{aa}) \\ \dfrac{\partial f_3}{\partial M^S_{Aa}} = -\dfrac{1}{\kappa^S}\, \beta^S_{aa}\, (\alpha\, \gamma^S_{aa}\, M^S_{aa} + (1-\alpha)\, \gamma^R_{aa}\, M^R_{aa}) \\ \dfrac{\partial f_3}{\partial M^S_{aa}} = \alpha\, \beta^S_{aa}\, \gamma^S_{aa}\, (1 - \dfrac{M^S}{\kappa^S}) - \dfrac{1}{\kappa^S}\, \beta^S_{aa}\, (\alpha\, \gamma^S_{aa}\, M^S_{aa} + (1-\alpha)\, \gamma^R_{aa}\, M^R_{aa}) - \mu \\ \dfrac{\partial f_3}{\partial M^R_{aa}} = \beta^S_{aa}\, (1-\alpha)\, \gamma^R_{aa}\, (1 - \dfrac{M^S}{\kappa^S}) \end{cases}$$

$$\begin{cases} \dfrac{\partial f_4}{\partial M^S_{AA}} = 0 \\ \dfrac{\partial f_4}{\partial M^S_{Aa}} = 0 \\ \dfrac{\partial f_4}{\partial M^S_{aa}} = \beta^R_{aa}\, (1-\alpha)\, \gamma^S_{aa}\, (1 - \dfrac{M^R}{\kappa^R}) \\ \dfrac{\partial f_4}{\partial M^R_{aa}} = \alpha\, \beta^R_{aa}\, \gamma^R_{aa}\, (1 - \dfrac{M^R}{\kappa^R}) - \dfrac{1}{\kappa^R}\, \beta^R_{aa}\, (\alpha\, \gamma^R_{aa}\, M^R_{aa} + (1-\alpha)\, \gamma^S_{aa}\, M^S_{aa}) - \mu \end{cases}$$

2.2.3.1 Extinction de la population

Nous cherchons l'équilibre correspondant à une extinction de la population (1. La disparition de l'ensemble des génotypes), *i.e.* pour lequel l'ensemble des conditions suivantes sont vérifiées :

$$\begin{cases} M_{AA}^S = 0 \\ M_{Aa}^S = 0 \\ M_{aa}^S = 0 \\ M_{aa}^R = 0 \\ \dfrac{dM_{AA}^S}{dt} = 0 \\ \dfrac{dM_{Aa}^S}{dt} = 0 \\ \dfrac{dM_{aa}^S}{dt} = 0 \\ \dfrac{dM_{aa}^R}{dt} = 0 \end{cases}$$

Le point d'équilibre correspondant est $E^0 = (0, 0, 0, 0)$

Étude de la stabilité de l'équilibre $E^0 = (0, 0, 0, 0)$

La matrice Jacobienne au point d'équilibre E^0 est la suivante :

$$J(E^0) = \begin{pmatrix} \beta_{AA}^S \, \gamma_{AA}^S \, \alpha - \mu & 0 & 0 & 0 \\ 0 & \beta_{Aa}^S \, \gamma_{Aa}^S \, \alpha - \mu & 0 & 0 \\ 0 & 0 & \beta_{aa}^S \, \gamma_{aa}^S \, \alpha - \mu & \beta_{aa}^S \, (1-\alpha) \gamma_{aa}^R \\ 0 & 0 & \beta_{aa}^R \, (1-\alpha) \gamma_{aa}^S & \beta_{aa}^R \, \gamma_{aa}^R \, \alpha - \mu \end{pmatrix}$$

En identifiant la première valeur propre $\lambda_1 = \beta_{AA}^S \, \gamma_{AA}^S \, \alpha - \mu$, nous en déduisons une première condition de stabilité N_1^0 de cet équilibre pour $\lambda_1 < 0$:

$$N_1^0 : \frac{\beta_{AA}^S \, \gamma_{AA}^S \, \alpha}{\mu} < 1$$

Si cette condition est vérifiée, elle satisfait également la condition de stabilité donnée par la seconde valeur propre : $N_2^0 : \lambda_2 = \beta_{Aa}^S \, \gamma_{Aa}^S \, \alpha - \mu < 0$ car $\beta_{AA}^S > \beta_{Aa}^S$ et $\gamma_{AA}^S > \gamma_{Aa}^S$

La matrice M nous permet d'étudier les conditions de stabilité associées aux troisième et quatrième valeurs propres :

$$M = \begin{pmatrix} \beta_{aa}^S \, \gamma_{aa}^S \, \alpha - \mu & \beta_{aa}^S \, (1-\alpha) \gamma_{aa}^R \\ \beta_{aa}^R \, (1-\alpha) \gamma_{aa}^S & \beta_{aa}^R \, \gamma_{aa}^R \, \alpha - \mu \end{pmatrix}$$

La trace et le déterminant de la matrice M correspondent respectivement à la somme et au produit des deux valeurs propres de M : $Tr(M) = \lambda_3 + \lambda_4$ et $Det(M) = \lambda_3 \times \lambda_4$. Ainsi, l'équilibre E^0 est stable si et seulement si ces deux valeurs propres sont négatives, donc si $Tr(M) < 0$ et $Det(M) > 0$:

$$Tr(M) = \alpha\,(\beta^S_{aa}\,\gamma^S_{aa} + \beta^R_{aa}\,\gamma^R_{aa}) - 2\,\mu$$

$$Tr(M) < 0 \Rightarrow \mu > \frac{\alpha}{2}\,(\beta^S_{aa}\,\gamma^S_{aa} + \beta^R_{aa}\,\gamma^R_{aa}) \tag{2.6}$$

$$Det(M) = \mu^2 - \alpha\,(\beta^S_{aa}\,\gamma^S_{aa} + \beta^R_{aa}\,\gamma^R_{aa})\,\mu + \beta^S_{aa}\,\gamma^S_{aa}\,\beta^R_{aa}\,\gamma^R_{aa}\,(2\,\alpha - 1)$$

$Det(M) = a\,\mu^2 + b\,\mu + c$ est un polynôme de degré 2, avec $-\frac{b}{2\,a} = \frac{\alpha}{2}\,(\beta^S_{aa}\,\gamma^S_{aa} + \beta^R_{aa}\,\gamma^R_{aa})$. En remplaçant $\frac{\alpha}{2}\,(\beta^S_{aa}\,\gamma^S_{aa} + \beta^R_{aa}\,\gamma^R_{aa})$ dans (2.6), nous obtenons $Tr(M) < 0 \Rightarrow \mu > -\frac{b}{2\,a}$. Pour respecter les deux conditions $Tr(M) < 0$ et $Det(M) > 0$, μ doit donc être supérieur à la plus grande des deux racines du polynôme $Det(M)$.

La seconde condition de stabilité de l'équilibre E^0 est donc la suivante :

$$N^0_3 : \mu > \tfrac{\alpha}{2}\,(\beta^S_{aa}\,\gamma^S_{aa} + \beta^R_{aa}\,\gamma^R_{aa}) + \tfrac{1}{2}\,\sqrt{\alpha^2\,(\beta^R_{aa}\,\gamma^R_{aa} + \beta^S_{aa}\,\gamma^S_{aa})^2 - 4\,(2\,\alpha - 1)\,\beta^S_{aa}\,\gamma^S_{aa}\,\beta^R_{aa}\,\gamma^R_{aa}}$$

L'équilibre $E^0 = (0, 0, 0, 0)$ est stable si

$$\mu > max(\beta^S_{AA}\,\gamma^S_{AA}\,\alpha\,;\ \frac{\alpha}{2}\,(\beta^S_{aa}\,\gamma^S_{aa} + \beta^R_{aa}\,\gamma^R_{aa}) + \frac{1}{2}\,\sqrt{\alpha^2\,(\beta^R_{aa}\,\gamma^R_{aa} + \beta^S_{aa}\,\gamma^S_{aa})^2 - 4\,(2\,\alpha - 1)\,\beta^S_{aa}\,\gamma^S_{aa}\,\beta^R_{aa}\,\gamma^R_{aa}}\,)$$

L'équilibre d'extinction existe donc toujours mais n'est stable que lorsque le taux de mortalité de la population d'agent pathogène est suffisamment élevé.

2.2.3.2 Disparition des individus hétérozygotes Aa

Nous cherchons l'équilibre pour lequel les individus de génotype Aa uniquement disparaissent (condition 2.), *i.e.* pour lequel l'ensemble des conditions suivantes sont vérifiées :

$$\begin{cases} M^S_{AA} > 0 \\ M^S_{aa} > 0 \\ M^R_{aa} > 0 \\ M^S_{Aa} = 0 \\ \dfrac{dM^S_{AA}}{dt} = 0 \\ \dfrac{dM^S_{Aa}}{dt} = 0 \\ \dfrac{dM^S_{aa}}{dt} = 0 \\ \dfrac{dM^R_{aa}}{dt} = 0 \end{cases}$$

Ce qui implique :

$$\begin{cases} \alpha\,\beta^S_{AA}\,\gamma^S_{AA}\,M^S_{AA}\,(1-\dfrac{M^S}{\kappa^S}) - \mu\,M^S_{AA} = 0 \\ \beta^S_{aa}\left(\alpha\,\gamma^S_{aa}\,M^S_{aa} + (1-\alpha)\,\gamma^R_{aa}\,M^R_{aa}\right)(1-\dfrac{M^S}{\kappa^S}) - \mu\,M^S_{aa} = 0 \\ \beta^R_{aa}\left(\alpha\,\gamma^R_{aa}\,M^R_{aa} + (1-\alpha)\,\gamma^S_{aa}\,M^S_{aa}\right)(1-\dfrac{M^R_{aa}}{\kappa^R}) - \mu\,M^R_{aa} = 0 \end{cases}$$

Nous obtenons l'expression de M^S en fonction des paramètres, et de M^R_{aa} en fonction de M^S_{aa} :

$$\begin{cases} M^S = \kappa^S\,(1-\dfrac{\mu}{\alpha\,\beta^S_{AA}\,\gamma^S_{AA}}) \\ M^R_{aa} = M^S_{aa}\,\alpha\,\dfrac{\beta^S_{AA}\,\gamma^S_{AA} - \beta^S_{aa}\,\gamma^S_{aa}}{(1-\alpha)\,\beta^S_{aa}\,\gamma^R_{aa}} \\ \beta^R_{aa}\left(\alpha\,\gamma^R_{aa}\,M^R_{aa} + (1-\alpha)\,\gamma^S_{aa}\,M^S_{aa}\right)(1-\dfrac{M^R_{aa}}{\kappa^R}) - \mu\,M^R_{aa} = 0 \end{cases}$$

En sachant $M^S_{AA} \neq 0$, $M^S_{aa} \neq 0$ et $M^R_{aa} \neq 0$, nous obtenons :

$$\begin{cases} M^S_{AA} = \kappa^S\,(1-\dfrac{\mu}{\alpha\,\beta^S_{AA}\,\gamma^S_{AA}}) - M^S_{aa} \\ M^R_{aa} = M^S_{aa}\,\alpha\,\dfrac{\beta^S_{AA}\,\gamma^S_{AA} - \beta^S_{aa}\,\gamma^S_{aa}}{(1-\alpha)\,\beta^S_{aa}\,\gamma^R_{aa}} \\ M^S_{aa} = (1-\alpha)\,\kappa^R\,\beta^S_{aa}\,(\dfrac{\gamma^R_{aa}}{\alpha\,(\beta^S_{AA}\,\gamma^S_{AA} - \beta^S_{aa}\,\gamma^S_{aa})} - \dfrac{\mu}{\beta^R_{aa}\,(\alpha^2\,\beta^S_{AA}\,\gamma^S_{AA} + \beta^S_{aa}\,\gamma^S_{aa}\,(1-2\,\alpha)}) \end{cases}$$

Et en remplaçant M^S_{aa} :

$$\begin{cases} M^S_{AA} = \kappa^S\,(1-\dfrac{\mu}{\alpha\,\beta^S_{AA}\,\gamma^S_{AA}}) - (1-\alpha)\,\kappa^R\,\beta^S_{aa}\,(\dfrac{\gamma^R_{aa}}{\alpha\,(\beta^S_{AA}\,\gamma^S_{AA} - \beta^S_{aa}\,\gamma^S_{aa})} - \dfrac{\mu}{\beta^R_{aa}\,(\alpha^2\,\beta^S_{AA}\,\gamma^S_{AA} + \beta^S_{aa}\,\gamma^S_{aa}\,(1-2\,\alpha)}) \\ M^R_{aa} = \kappa^R\,(1-\dfrac{\mu\,\alpha\,(\beta^S_{AA}\,\gamma^S_{AA} - \beta^S_{aa}\,\gamma^S_{aa})}{\beta^R_{aa}\,\gamma^R_{aa}\,(\alpha^2\,\beta^S_{AA}\,\gamma^S_{AA} + \beta^S_{aa}\,\gamma^S_{aa}\,(1-2\,\alpha))}) \\ M^S_{aa} = (1-\alpha)\,\kappa^R\,\beta^S_{aa}\,(\dfrac{\gamma^R_{aa}}{\alpha\,(\beta^S_{AA}\,\gamma^S_{AA} - \beta^S_{aa}\,\gamma^S_{aa})} - \dfrac{\mu}{\beta^R_{aa}\,(\alpha^2\,\beta^S_{AA}\,\gamma^S_{AA} + \beta^S_{aa}\,\gamma^S_{aa}\,(1-2\,\alpha)}) \end{cases}$$

Pour simplifier ces conditions, comme $\beta^S_{AA} > \beta^S_{Aa} > \beta^S_{aa}$ et $\gamma^S_{AA} > \gamma^S_{Aa} > \gamma^S_{aa}$, nous posons :

$\beta^S_{AA} = \beta^S$; $\beta^S_{aa} = \epsilon^\beta\,\beta^S$; $\beta^S_{Aa} = (\epsilon^\beta + (1-\epsilon^\beta)\,\psi^\beta)\,\beta^S$ et $\beta^R_{aa} = \beta^R$

$\gamma^S_{AA} = \gamma^S$; $\gamma^S_{aa} = \epsilon^\gamma\,\gamma^S$; $\gamma^S_{Aa} = (\epsilon^\gamma + (1-\epsilon^\gamma)\,\psi^\gamma)\,\gamma^S$ et $\gamma^R_{aa} = \gamma^R$

avec ϵ le paramètre du coût sélectif et ψ le paramètre de pondération de ce coût sélectif pour les hétérozygotes, tous les deux entre 0 et 1.

En posant de nouvelles hypothèses : $\epsilon^{\beta} = \epsilon^{\gamma} = \epsilon$ et $\psi^{\beta} = \psi^{\gamma} = \psi$ nous obtenons :
$\beta^S_{AA} = \beta^S$; $\beta^S_{aa} = \epsilon\,\beta^S$; $\beta^S_{Aa} = (\epsilon + (1-\epsilon)\,\psi)\,\beta^S$ et $\beta^R_{aa} = \beta^R$
$\gamma^S_{AA} = \gamma^S$; $\gamma^S_{aa} = \epsilon\,\gamma^S$; $\gamma^S_{Aa} = (\epsilon + (1-\epsilon)\,\psi)\,\gamma^S$ et $\gamma^R_{aa} = \gamma^R$

Avec ces nouvelles hypothèses, nous obtenons l'équilibre suivant $E^1 = (M^S_{AA}, 0, M^S_{aa}, M^R_{aa})$ avec :

$$\begin{cases} M^S_{AA} = \kappa^S\,(1 - \dfrac{\mu}{\alpha\,\beta^S\,\gamma^S}) - \dfrac{(1-\alpha)\,\kappa^R\,\epsilon}{\gamma^S}\,(\dfrac{\gamma^R}{\alpha\,(1-\epsilon)} - \dfrac{\mu}{\beta^R\,(\alpha^2 + \epsilon^2\,(1-2\,\alpha))}) \\ M^S_{aa} = \dfrac{(1-\alpha)\,\kappa^R\,\epsilon}{\gamma^S}\,(\dfrac{\gamma^R}{\alpha\,(1-\epsilon)} - \dfrac{\mu}{\beta^R\,(\alpha^2 + \epsilon^2\,(1-2\,\alpha))}) \\ M^R_{aa} = \kappa^R\,(1 - \dfrac{\mu\,\alpha\,(1-\epsilon)}{\beta^R\,\gamma^R\,(\alpha^2 + \epsilon^2\,(1-2\,\alpha))} \end{cases} \tag{2.7}$$

La construction de la matrice Jacobienne ne nous permet pas de déterminer analytiquement les conditions de stabilité de cet équilibre et nécessiterait une approximation numérique.

2.2.3.3 Disparition des individus homozygotes AA

Nous cherchons l'équilibre pour lequel les individus de génotype AA uniquement disparaissent (condition 3.), *i.e.* pour lequel l'ensemble des conditions suivantes sont vérifiées :

$$\begin{cases} M^S_{Aa} > 0 \\ M^S_{aa} > 0 \\ M^R_{aa} > 0 \\ M^S_{AA} = 0 \\ \dfrac{dM^S_{AA}}{dt} = 0 \\ \dfrac{dM^S_{Aa}}{dt} = 0 \\ \dfrac{dM^S_{aa}}{dt} = 0 \\ \dfrac{dM^R_{aa}}{dt} = 0 \end{cases}$$

Ce qui implique :

$$\begin{cases} \alpha\,\beta^S_{Aa}\,\gamma^S_{Aa}\,M^S_{Aa}\,(1-\dfrac{M^S}{\kappa^S})-\mu\,M^S_{Aa}=0 \\ \beta^S_{aa}\left(\alpha\,\gamma^S_{aa}\,M^S_{aa}+(1-\alpha)\,\gamma^R_{aa}\,M^R_{aa}\right)(1-\dfrac{M^S}{\kappa^S})-\mu\,M^S_{aa}=0 \\ \beta^R_{aa}\left(\alpha\,\gamma^R_{aa}\,M^R_{aa}+(1-\alpha)\,\gamma^S_{aa}\,M^S_{aa}\right)(1-\dfrac{M^R_{aa}}{\kappa^R})-\mu\,M^R_{aa}=0 \end{cases}$$

Nous obtenons l'expression de M^S en fonction des paramètres, et de M^R_{aa} en fonction de M^S_{aa} :

$$\begin{cases} M^S=\kappa^S\,(1-\dfrac{\mu}{\alpha\,\beta^S_{Aa}\,\gamma^S_{Aa}}) \\ M^R_{aa}=M^S_{aa}\,\alpha\,\dfrac{\beta^S_{Aa}\,\gamma^S_{Aa}\;-\;\beta^S_{aa}\,\gamma^S_{aa}}{(1-\alpha)\,\beta^S_{aa}\,\gamma^R_{aa}} \\ \beta^R_{aa}\left(\alpha\,\gamma^R_{aa}\,M^R_{aa}+(1-\alpha)\,\gamma^S_{aa}\,M^S_{aa}\right)(1-\dfrac{M^R_{aa}}{\kappa^R})-\mu\,M^R_{aa}=0 \end{cases}$$

En sachant $M^S_{Aa}\neq 0$, $M^S_{aa}\neq 0$ et $M^R_{aa}\neq 0$:

$$\begin{cases} M^S_{Aa}=\kappa^S\,(1-\dfrac{\mu}{\alpha\,\beta^S_{Aa}\,\gamma^S_{Aa}})-M^S_{aa} \\ M^R_{aa}=M^S_{aa}\,\alpha\,\dfrac{\beta^S_{Aa}\,\gamma^S_{Aa}\;-\;\beta^S_{aa}\,\gamma^S_{aa}}{(1-\alpha)\,\beta^S_{aa}\,\gamma^R_{aa}} \\ M^S_{aa}=(1-\alpha)\,\kappa^R\,\beta^S_{aa}\,(\dfrac{\gamma^R_{aa}}{\alpha\,(\beta^S_{Aa}\,\gamma^S_{Aa}-\beta^S_{aa}\,\gamma^S_{aa})}\;-\;\dfrac{\mu}{\beta^R_{aa}\,(\alpha^2\,\beta^S_{Aa}\,\gamma^S_{Aa}+\beta^S_{aa}\,\gamma^S_{aa}\,(1-2\,\alpha)}) \end{cases}$$

Et en remplaçant M^S_{aa}, nous obtenons l'équilibre suivant $E^2=(M^S_{AA},0,M^S_{aa},M^R_{aa})$ avec :

$$\begin{cases} M^S_{Aa}=\kappa^S\,(1-\dfrac{\mu}{\alpha\,\beta^S_{Aa}\,\gamma^S_{Aa}}) \\ \qquad\qquad -(1-\alpha)\,\kappa^R\,\beta^S_{aa}\,(\dfrac{\gamma^R_{aa}}{\alpha\,(\beta^S_{Aa}\,\gamma^S_{Aa}-\beta^S_{aa}\,\gamma^S_{aa})}\;-\;\dfrac{\mu}{\beta^R_{aa}\,(\alpha^2\,\beta^S_{Aa}\,\gamma^S_{Aa}+\beta^S_{aa}\,\gamma^S_{aa}\,(1-2\,\alpha)}) \\ M^R_{aa}=\kappa^R\,(1-\dfrac{\mu\,\alpha\,(\beta^S_{Aa}\,\gamma^S_{Aa}\;-\;\beta^S_{aa}\,\gamma^S_{aa})}{\beta^R_{aa}\,\gamma^R_{aa}\,(\alpha^2\,\beta^S_{Aa}\,\gamma^S_{Aa}+\beta^S_{aa}\,\gamma^S_{aa}\,(1-2\,\alpha))} \\ M^S_{aa}=(1-\alpha)\,\kappa^R\,\beta^S_{aa}\,(\dfrac{\gamma^R_{aa}}{\alpha\,(\beta^S_{Aa}\,\gamma^S_{Aa}-\beta^S_{aa}\,\gamma^S_{aa})}\;-\;\dfrac{\mu}{\beta^R_{aa}\,(\alpha^2\,\beta^S_{Aa}\,\gamma^S_{Aa}+\beta^S_{aa}\,\gamma^S_{aa}\,(1-2\,\alpha)}) \end{cases} \tag{2.8}$$

De même que dans la partie 2.2.3.2, la construction de la matrice Jacobienne ne nous

permet pas de déterminer analytiquement les conditions de stabilité de cet équilibre et nécessiterait une approximation numérique.

2.2.3.4 Fixation de l'allèle A

Nous cherchons l'équilibre correspondant à une fixation de l'allèle A dans la population (condition 4. La disparition des génotypes Aa, et aa sur S et R), *i.e.* pour lequel l'ensemble des conditions suivantes sont vérifiées :

$$\begin{cases} M^S_{AA} > 0 \\ M^S_{Aa} = 0 \\ M^S_{aa} = 0 \\ M^R_{aa} = 0 \\ \frac{dM^S_{AA}}{dt} = 0 \\ \frac{dM^S_{Aa}}{dt} = 0 \\ \frac{dM^S_{aa}}{dt} = 0 \\ \frac{dM^R_{aa}}{dt} = 0 \end{cases}$$

Ce qui implique :

$$M^S_{AA}\,(\beta^S_{AA}\,\gamma^S_{AA}\,\alpha\,(1-\frac{M^S_{AA}}{\kappa^S})-\mu)=0$$

$M^S_{AA} > 0$, donc le point d'équilibre correspondant est $E^3 = (\kappa^S\,(1-\frac{\mu}{\alpha\,\beta^S_{AA}\,\gamma^S_{AA}}),0,0,0)$ qui existe si $\mu < \beta^S_{AA}\,\gamma^S_{AA}\,\alpha$

Étude de la stabilité de l'équilibre $E^3 = (\kappa^S\,(1-\frac{\mu}{\alpha\,\beta^S_{AA}\,\gamma^S_{AA}}),0,0,0)$

La matrice Jacobienne au point d'équilibre E^3 est la suivante :

$$J(E^3) = \begin{pmatrix} \mu-\beta^S_{AA}\,\gamma^S_{AA}\,\alpha & \mu-\beta^S_{AA}\,\gamma^S_{AA}\,\alpha & \mu-\beta^S_{AA}\,\gamma^S_{AA}\,\alpha & 0 \\ 0 & \mu\,(\frac{\beta^S_{Aa}\,\gamma^S_{Aa}}{\beta^S_{AA}\,\gamma^S_{AA}}-1) & 0 & 0 \\ 0 & 0 & \mu\,(\frac{\beta^S_{aa}\,\gamma^S_{aa}}{\beta^S_{AA}\,\gamma^S_{AA}}-1) & \mu\,(\frac{\beta^S_{aa}\,(1-\alpha)\,\gamma^R_{aa}}{\beta^S_{AA}\,\gamma^S_{AA}\,\alpha}) \\ 0 & 0 & \beta^R_{aa}\,(1-\alpha)\gamma^S_{aa} & \beta^R_{aa}\,\gamma^R_{aa}\,\alpha-\mu \end{pmatrix}$$

En identifiant la première valeur propre $\lambda_1 = \mu - \beta^S_{AA}\,\gamma^S_{AA}\,\alpha$, nous en déduisons une première condition de stabilité qui est la condition d'existence de cet équilibre donnée précédemment :

$N_1^3 : \mu < \beta_{AA}^S \, \gamma_{AA}^S \, \alpha$

La seconde valeur propre $\lambda_2 = \mu(\frac{\beta_{Aa}^S \, \gamma_{Aa}^S}{\beta_{AA}^S \, \gamma_{AA}^S} - 1)$ nous donne une seconde condition de stabilité :

$N_2^3 : \mu \, (\frac{\beta_{Aa}^S \, \gamma_{Aa}^S}{\beta_{AA}^S \, \gamma_{AA}^S} - 1) < 0$

Cette condition est satisfaite pour $\mu > 0$ car $\beta_{AA}^S > \beta_{Aa}^S$ et $\gamma_{AA}^S > \gamma_{Aa}^S$.
La matrice M nous permet d'étudier les conditions de stabilité associées aux troisième et quatrième valeurs propres :

$$N = \begin{pmatrix} \mu \, (\frac{\beta_{aa}^S \, \gamma_{aa}^S}{\beta_{AA}^S \, \gamma_{AA}^S} - 1) & \mu \, (\frac{\beta_{aa}^S \, (1-\alpha) \, \gamma_{aa}^R}{\beta_{AA}^S \, \gamma_{AA}^S \, \alpha}) \\ \beta_{aa}^R \, (1-\alpha) \gamma_{aa}^S & \beta_{aa}^R \, \gamma_{aa}^R \, \alpha - \mu \end{pmatrix}$$

L'équilibre E^3 est stable si $Tr(N) < 0$ et $Det(N) > 0$:

$$Tr(N) = \mu \, (\frac{\beta_{aa}^S \, \gamma_{aa}^S}{\beta_{AA}^S \, \gamma_{AA}^S) - 2) + \alpha \, \beta_{aa}^R \, \gamma_{aa}^R$$

$$Tr(N) < 0 \Rightarrow \mu > \frac{\beta_{aa}^R \, \gamma_{aa}^R \, \alpha \, \beta_{AA}^S \, \gamma_{AA}^S}{2 \, \beta_{AA}^S \, \gamma_{AA}^S - \beta_{aa}^S \, \gamma_{aa}^S}$$

Remarque : $\beta_{AA}^S \, \gamma_{AA}^S > \beta_{aa}^S \, \gamma_{aa}^S$, donc $\frac{\beta_{aa}^R \, \gamma_{aa}^R \, \alpha \, \beta_{AA}^S \, \gamma_{AA}^S}{2 \, \beta_{AA}^S \, \gamma_{AA}^S - \beta_{aa}^S \, \gamma_{aa}^S} < \frac{\beta_{aa}^R \, \gamma_{aa}^R \, \alpha \, \beta_{AA}^S \, \gamma_{AA}^S}{2 \, \beta_{AA}^S \, \gamma_{AA}^S - \beta_{AA}^S \, \gamma_{AA}^S} = \beta_{aa}^R \, \gamma_{aa}^R \, \alpha$.

Une condition suffisante pour observer $Tr(N) < 0$ est donc $\mu > \beta_{aa}^R \, \gamma_{aa}^R \, \alpha$.

$$Det(N) = \mu \, (\frac{\beta_{aa}^S \, \gamma_{aa}^S}{\beta_{AA}^S \, \gamma_{AA}^S} - 1) \left(\beta_{aa}^R \, \gamma_{aa}^R \, \alpha - \mu\right) - \mu \, (\frac{\beta_{aa}^S \, (1-\alpha) \, \gamma_{aa}^R}{\beta_{AA}^S \, \gamma_{AA}^S \, \alpha}) \beta_{aa}^R \, (1-\alpha) \gamma_{aa}^S$$

$$Det(N) = \mu \, (\mu \, (1 - \frac{\beta_{aa}^S \, \gamma_{aa}^S}{\beta_{AA}^S \, \gamma_{AA}^S}) - \alpha \, \beta_{aa}^R \, \gamma_{aa}^R - \frac{\beta_{aa}^R \, \gamma_{aa}^R \, \beta_{aa}^S \, \gamma_{aa}^S \, [(1-\alpha)^2 - \alpha^2]}{\alpha \, \beta_{AA}^S \, \gamma_{AA}^S})$$

$$\begin{aligned} Det(N) > 0 &\Rightarrow \mu \, (\mu \, (1 - \frac{\beta_{aa}^S \, \gamma_{aa}^S}{\beta_{AA}^S \, \gamma_{AA}^S}) - \alpha \, \beta_{aa}^R \, \gamma_{aa}^R - \frac{\beta_{aa}^R \, \gamma_{aa}^R \, \beta_{aa}^S \, \gamma_{aa}^S \, [(1-\alpha)^2 - \alpha^2]}{\alpha \, \beta_{AA}^S \, \gamma_{AA}^S}) > 0 \\ &\Rightarrow \mu \, (1 - \frac{\beta_{aa}^S \, \gamma_{aa}^S}{\beta_{AA}^S \, \gamma_{AA}^S}) - \alpha \, \beta_{aa}^R \, \gamma_{aa}^R - \frac{\beta_{aa}^R \, \gamma_{aa}^R \, \beta_{aa}^S \, \gamma_{aa}^S \, [(1-\alpha)^2 - \alpha^2]}{\alpha \, \beta_{AA}^S \, \gamma_{AA}^S} > 0 \\ &\Rightarrow \mu > \frac{\beta_{aa}^R \, \gamma_{aa}^R \, \beta_{AA}^S \, \gamma_{AA}^S \, \alpha^2 - \beta_{aa}^R \, \gamma_{aa}^R \, \beta_{aa}^S \, \gamma_{aa}^S \alpha^2 + \beta_{aa}^R \, \gamma_{aa}^R \, \beta_{aa}^S \, \gamma_{aa}^S \, (1-\alpha)^2}{\alpha \, (\beta_{AA}^S \, \gamma_{AA}^S - \beta_{aa}^S \, \gamma_{aa}^S)} \\ &= \beta_{aa}^R \, \gamma_{aa}^R \alpha \frac{\alpha(\beta_{AA}^S \, \gamma_{AA}^S - \beta_{aa}^S \, \gamma_{aa}^S) \; + \; \beta_{aa}^S \, \gamma_{aa}^S \, \frac{1}{\alpha}(1-\alpha)^2}{\alpha(\, \beta_{AA}^S \, \gamma_{AA}^S - \beta_{aa}^S \, \gamma_{aa}^S)} \end{aligned}$$

Remarque : Comme $\beta_{AA}^S \, \gamma_{AA}^S - \beta_{aa}^S \, \gamma_{aa}^S > 0$, le quotient est plus grand que 1 et $\mu > \beta_{aa}^R \, \gamma_{aa}^R \, \alpha$ est une condition nécessaire pour observer $Det(N) > 0$. $Det(N) > 0 \Rightarrow Tr(N) < 0$, donc $Det(N) > 0$ est la seule condition à vérifier.

Il existe donc une unique condition de stabilité de l'équilibre E^3 :

$$N_3^3 : \mu > \frac{\beta_{aa}^R\,\gamma_{aa}^R\,\beta_{AA}^S\,\gamma_{AA}^S\,\alpha^2+\beta_{aa}^R\,\gamma_{aa}^R\,\beta_{aa}^S\,\gamma_{aa}^S\,(1-2\,\alpha)}{\alpha\,(\beta_{AA}^S\,\gamma_{AA}^S-\beta_{aa}^S\,\gamma_{aa}^S)} = \beta_{aa}^R\,\gamma_{aa}^R\alpha\frac{\alpha(\beta_{AA}^S\,\gamma_{AA}^S-\beta_{aa}^S\,\gamma_{aa}^S)+\beta_{aa}^S\,\gamma_{aa}^S\,\frac{1}{\alpha}(1-\alpha)^2}{\alpha(\,\beta_{AA}^S\,\gamma_{AA}^S-\beta_{aa}^S\,\gamma_{aa}^S)}$$

L'équilibre $E^3 = (\kappa^S\,(1-\frac{\mu}{\alpha\,\beta_{AA}^S\,\gamma_{AA}^S}), 0, 0, 0)$ existe si :

$$\mu < \beta_{AA}^S\,\gamma_{AA}^S\,\alpha$$

et est stable si :

$$\frac{\beta_{aa}^R\,\gamma_{aa}^R\,\beta_{AA}^S\,\gamma_{AA}^S\,\alpha^2+\beta_{aa}^R\,\gamma_{aa}^R\,\beta_{aa}^S\,\gamma_{aa}^S\,(1-2\,\alpha)}{\alpha\,(\beta_{AA}^S\,\gamma_{AA}^S-\beta_{aa}^S\,\gamma_{aa}^S)} < \mu$$

En posant les mêmes hypothèses que précédemment :
$\beta_{AA}^S = \beta^S$; $\beta_{aa}^S = \epsilon\,\beta^S$; $\beta_{Aa}^S = (\epsilon + (1-\epsilon)\,\psi)\,\beta^S$ et $\beta_{aa}^R = \beta^R$
$\gamma_{AA}^S = \gamma^S$; $\gamma_{aa}^S = \epsilon\,\gamma^S$; $\gamma_{Aa}^S = (\epsilon + (1-\epsilon)\,\psi)\,\gamma^S$ et $\gamma_{aa}^R = \gamma^R$

Avec ces nouvelles hypothèses uniquement, l'équilibre $E^3 = (\kappa^S\,(1-\frac{\mu}{\alpha\,\beta^S\,\gamma^S}), 0, 0, 0)$ existe si :

$$\mu < \beta^S\,\gamma^S\,\alpha$$

et est stable si :

$$\frac{\beta^R\,\gamma^R\,(\alpha^2+\epsilon^2\,(1-2\,\alpha))}{\alpha\,(1-\epsilon^2)} < \mu$$

Pour que ces dernières conditions sur μ soient compatibles, les paramètres doivent vérifier :

$$\beta^S\,\gamma^S > \beta^R\,\gamma^R\,\frac{(\alpha^2+\epsilon^2\,(1-2\,\alpha))}{\alpha^2\,(1-\epsilon^2)} = \beta^R\,\gamma^R\frac{\alpha^2(1-\epsilon^2)+\epsilon^2\,(1-\alpha)^2}{\alpha^2\,(1-\epsilon^2)}$$

Cette fraction est supérieure à 1, il est donc nécessaire que $\beta^S\gamma^S > \beta^R\gamma^R$ pour observer une fixation de l'allèle A.

L'allèle avirulent A ne peut donc se fixer dans la population que si le taux de mortalité est suffisamment grand (condition d'existence de l'équilibre), mais n'est stable que si le taux de mortalité ne dépasse pas le seuil de stabilité de l'équilibre d'extinction de la population. De plus, la fixation de l'allèle A n'est possible que si les valeurs sélectives des individus sur le compartiment R sont inférieures aux valeurs sélectives sur S.

2.2.3.5 Fixation de l'allèle a

Nous cherchons les équilibres pour lesquels l'allèle a se fixe dans la population (correspondant à l'une des trois conditions 5., 6., ou 7.) *i.e.* pour lequel l'ensemble des conditions suivantes sont vérifiées :

$$\begin{cases} M_{aa}^S > 0 \, et/ou \, M_{aa}^R > 0 \\ M_{AA}^S = 0 \\ M_{Aa}^S = 0 \\ \dfrac{dM_{AA}^S}{dt} = 0 \\ \dfrac{dM_{Aa}^S}{dt} = 0 \\ \dfrac{dM_{aa}^S}{dt} = 0 \\ \dfrac{dM_{aa}^R}{dt} = 0 \end{cases} \tag{2.9}$$

Ce qui implique :

$$\begin{cases} \beta_{aa}^S \left(\alpha \, \gamma_{aa}^S \, M_{aa}^S + (1-\alpha) \, \gamma_{aa}^R \, M_{aa}^R\right) (1 - \dfrac{M_{aa}^S}{\kappa^S}) - \mu \, M_{aa}^S = 0 \\ \beta_{aa}^R \left(\alpha \, \gamma_{aa}^R \, M_{aa}^R + (1-\alpha) \, \gamma_{aa}^S \, M_{aa}^S\right) (1 - \dfrac{M_{aa}^R}{\kappa^R}) - \mu \, M_{aa}^R = 0 \end{cases}$$

Nous exprimons M_{aa}^S en fonction de M_{aa}^R :

$$\begin{cases} \beta_{aa}^S \left(\alpha \, \gamma_{aa}^S \, M_{aa}^S + (1-\alpha) \, \gamma_{aa}^R \, M_{aa}^R\right) (1 - \dfrac{M_{aa}^S}{\kappa^S}) - \mu \, M_{aa}^S = 0 \\ M_{aa}^R \, \dfrac{1}{(1-\alpha) \, \gamma_{aa}^S} \, (\dfrac{\mu \, \kappa^R}{\beta_{aa}^R \, (\kappa^R - M_{aa}^R)} - \gamma_{aa}^R \, \alpha) = M_{aa}^S \end{cases}$$

En sachant $M_{aa}^S \neq 0$ et $M_{aa}^R \neq 0$, nous obtenons un polynôme de degré 3 sur M_{aa}^R :

$$\begin{cases} a_3 \, (M_{aa}^R)^3 + a_2 \, (M_{aa}^R)^2 + a_1 \, M_{aa}^R + a_0 = 0 \\ a_0 = \kappa^{R^2} \beta_{aa}^R \, (\, \beta_{aa}^S \, \alpha \, \mu \, \kappa^S \, (1-\alpha) \, \gamma_{aa}^S \; - \; \mu^2 \, \kappa^S \, (1-\alpha) \; + \; \beta_{aa}^S \, \gamma_{aa}^R \, \kappa^S \, \gamma_{aa}^S \, \beta_{aa}^R \, (1 - 3\alpha + 2\alpha^2) \\ \qquad + \; \mu \, \gamma_{aa}^R \, \alpha \, \kappa^S \, (1-\alpha) \, \beta_{aa}^R \,) \\ a_1 = \kappa^R \, (\, \gamma_{aa}^{R^2} \, \beta_{aa}^S \, \beta_{aa}^{R^2} \, \kappa^R \, \alpha(1-2\alpha) \; - \; 2 \, \mu \, \gamma_{aa}^R \, \alpha \, \kappa^S \, (1-\alpha) \, \beta_{aa}^{R^2} \; + \; 2 \, \beta_{aa}^S \, \gamma_{aa}^R \, \kappa^S \, \gamma_{aa}^S \, \beta_{aa}^{R^2} \, (-1 + 3\alpha - 2\alpha^2) \\ \qquad + \; \mu^2 \, \kappa^S \, (1-\alpha) \, \beta_{aa}^R \; + \; \mu \, \kappa^R \, \beta_{aa}^S \, \gamma_{aa}^R \, \beta_{aa}^R \, (-1 + \alpha(\alpha+2)) \; - \; \mu^2 \, \kappa^R \, \beta_{aa}^S \, \alpha \; - \; \beta_{aa}^S \, \alpha \, \mu \, \kappa^S \, (1-\alpha) \, \gamma_{aa}^S \, \beta_{aa}^R \,) \\ a_2 = \beta_{aa}^R \, \gamma_{aa}^R \, (\mu \, \kappa^R \, \beta_{aa}^S \, (1 - 2\alpha - \alpha^2) \; + \; \beta_{aa}^S \, \kappa^S \, \gamma_{aa}^S \, \beta_{aa}^R \, (1 - 3\alpha + 2\alpha^2) \\ \qquad + \; 2 \, \kappa^R \, \gamma_{aa}^R \, \beta_{aa}^S \, \beta_{aa}^R \, \alpha(2\alpha - 1) \; + \; \mu \, \alpha \, \kappa^S \, (1-\alpha) \, \beta_{aa}^R \,) \\ a_3 = \alpha \, \gamma_{aa}^{R^2} \, \beta_{aa}^S \, \beta_{aa}^{R^2} \, (1 - 2\alpha) \end{cases} \tag{2.10}$$

Ce polynôme permet d'obtenir certaines informations sur les conditions d'existence de solutions positives, selon le signe des différents coefficients. En particulier en posant l'hypothèse $\alpha > 0,5$, alors nous obtenons $a_3 < 0$ et :

a_3	a_2	a_1	a_0	**Solutions positives**
-	+	+	-	0 ou 2 Solutions
-	+	-	-	0 ou 2 Solutions
-	-	-	-	0 Solution
-	-	+	+	1 Solution
-	-	-	+	1 Solution
-	+	-	+	1 ou 3 Solutions
-	+	+	+	1 Solution
-	-	+	-	0 ou 2 Solutions

Pour $\alpha > 0,5$, il y a donc entre 0 et 3 solutions possibles correspondant à la fixation de l'allèle a dans la population. Le modèle analytique ne nous permet pas de connaître ces équilibres et d'étudier leur stabilité, ni de déterminer le nombre d'équilibres dans le cas où $\alpha \leq 0,5$. Ces éléments nécessiteraient une résolution par approximation numérique.

2.2.3.6 Survie des individus hétérozygotes Aa uniquement

Nous cherchons l'équilibre pour lequel les individus de génotype Aa sont les seuls survivants (condition 8). La disparition des génotypes AA et aa sur S et R, *i.e.* pour lequel l'ensemble des conditions suivantes sont vérifiées :

$$\begin{cases} M_{Aa}^S > 0 \\ M_{AA}^S = 0 \\ M_{aa}^S = 0 \\ M_{aa}^R = 0 \\ \dfrac{dM_{AA}^S}{dt} = 0 \\ \dfrac{dM_{Aa}^S}{dt} = 0 \\ \dfrac{dM_{aa}^S}{dt} = 0 \\ \dfrac{dM_{aa}^R}{dt} = 0 \end{cases}$$

Ce qui implique :

$$M_{Aa}^S \left(\beta_{Aa}^S \, \gamma_{Aa}^S \, \alpha \left(1 - \frac{M_{Aa}^S}{\kappa^S}\right) - \mu\right) = 0$$

En sachant $M_{Aa}^S > 0$, le point d'équilibre correspondant est le point $E^4 = (0, \kappa^S \left(1 - \frac{\mu}{\alpha \, \beta_{Aa}^S \, \gamma_{Aa}^S}\right), 0, 0)$ qui existe si $\mu < \beta_{Aa}^S \, \gamma_{Aa}^S \, \alpha$

Étude de la stabilité de l'équilibre $E^4 = (0, \kappa^S\,(1 - \frac{\mu}{\alpha\,\beta^S_{Aa}\,\gamma^S_{Aa}}), 0, 0)$

La matrice Jacobienne au point d'équilibre E^4 est la suivante :

$$J(E^4) = \begin{pmatrix} \mu - \beta^S_{Aa}\,\gamma^S_{Aa}\,\alpha & \mu - \beta^S_{Aa}\,\gamma^S_{Aa}\,\alpha & \mu - \beta^S_{Aa}\,\gamma^S_{Aa}\,\alpha & 0 \\ 0 & \mu\,(\frac{\beta^S_{AA}\,\gamma^S_{AA}}{\beta^S_{Aa}\,\gamma^S_{Aa}} - 1) & 0 & 0 \\ 0 & 0 & \mu\,(\frac{\beta^S_{aa}\,\gamma^S_{aa}}{\beta^S_{Aa}\,\gamma^S_{Aa}} - 1) & \mu\,(\frac{\beta^S_{aa}\,(1-\alpha)\,\gamma^R_{aa}}{\beta^S_{Aa}\,\gamma^S_{Aa}\,\alpha}) \\ 0 & 0 & \beta^R_{aa}\,(1-\alpha)\gamma^S_{aa} & \beta^R_{aa}\,\gamma^R_{aa}\,\alpha - \mu \end{pmatrix}$$

En identifiant la première valeur propre $\lambda_1 = \mu - \beta^S_{Aa}\gamma^S_{Aa}\alpha$, nous en déduisons une première condition de stabilité qui est également la condition d'existence de cet équilibre :

$N^4_1 : \mu < \beta^S_{Aa}\,\gamma^S_{Aa}\,\alpha$

La seconde valeur propre $\lambda_2 = \mu(\frac{\beta^S_{AA}\,\gamma^S_{AA}}{\beta^S_{Aa}\,\gamma^S_{Aa}} - 1)$ nous donne une seconde condition de stabilité :

$N^4_2 : \mu\,(\frac{\beta^S_{AA}\,\gamma^S_{AA}}{\beta^S_{Aa}\,\gamma^S_{Aa}} - 1) < 0$

Cette condition n'est satisfaite que si $\mu < 0$ car $\beta^S_{AA} > \beta^S_{Aa}$ et $\gamma^S_{AA} > \gamma^S_{Aa}$, or $\mu \geq 0$

L'équilibre $E^4 = (0, \kappa^S\,(1 - \frac{\mu}{\alpha\,\beta^S_{Aa}\,\gamma^S_{Aa}}), 0, 0)$ existe si :

$$\mu < \beta^S_{Aa}\,\gamma^S_{Aa}\,\alpha$$

Et n'est jamais stable.
Les individus hétérozygotes Aa ne sont donc jamais les seuls survivants du système à la fin de la saison épidémique clonale.

2.2.3.7 Conditions restantes

Les conditions restantes permettent d'évaluer rapidement les conditions d'existence des équilibres correspondants et de conclure sur l'absence d'existence de ces points d'équilibre (Tableau 2.2).

La condition 10. (disparition des génotypes AA et aa sur R) n'est jamais remplie pour $\beta^S_{Aa}\,\gamma^S_{Aa} > \beta^S_{aa}\,\gamma^S_{aa}$ et ne correspond donc à aucun équilibre.

La condition 12. (disparition des génotypes Aa et aa sur R) n'est jamais remplie pour $\beta^S_{AA}\,\gamma^S_{AA} > \beta^S_{aa}\,\gamma^S_{aa}$ et ne correspond donc à aucun équilibre.

Les conditions suivantes ne sont jamais remplies pour $\mu \neq 0$ et ne correspondent donc à aucun équilibre :

- La condition 9. (disparition des génotypes AA et aa sur S)
- La condition 11. (disparition des génotypes Aa et aa sur S)

Les conditions suivantes ne sont jamais remplies pour $\beta^S_{AA}\,\gamma^S_{AA} > \beta^S_{Aa}\,\gamma^S_{Aa}$ et ne correspondent donc à aucun équilibre :

- La condition 13. (disparition des génotypes aa sur S et R)
- La condition 14. (disparition des génotypes aa sur S)
- La condition 15. (disparition des génotypes aa sur R)
- La condition 16. (survie de l'ensemble des génotypes)

TABLEAU 2.2 – Tableau regroupant l'ensemble des conditions étudiées, les conditions d'existence des équilibres associés, les équilibres, et leurs conditions de stabilité lorsqu'elles ont pu être calculées.

Equilibre	**Condition d'existence**	**Equilibre**	**Condition de stabilité**
1. Disparition des AA, Aa, aaS et aaR → Extinction	Existe toujours	$E^0 = (0,0,0,0)$	$\mu > max(\beta^S_{AA}\,\gamma^S_{AA}\,\alpha\,;\ \frac{\alpha}{2}\,(\beta^S_{aa}\,\gamma^S_{aa} + \beta^R_{aa}\,\gamma^R_{aa}) + \frac{1}{2}\,\sqrt{\alpha^2\,(\beta^R_{aa}\,\gamma^R_{aa} + \beta^S_{aa}\,\gamma^S_{aa})^2 - 4\,(2\,\alpha - 1)\,\beta^S_{aa}\,\gamma^S_{aa}\,\beta^R_{aa}\,\gamma^R_{aa}}\ )$
2. Disparition des Aa	Inconnue	Équation (2.7)	Jacobienne non étudiée
3. Disparition des AA	Inconnue	Équation (2.7)	Jacobienne non étudiée
4. Disparition des Aa, aaS et aaR → Fixation de A	$\mu < \beta^S_{AA}\,\gamma^S_{AA}\,\alpha$	$E^3 = (\kappa^S\,(1 - \frac{\mu}{\alpha\,\beta^S_{AA}\,\gamma^S_{AA}}), 0, 0, 0)$	$\frac{\beta^R_{aa}\,\gamma^R_{aa}\,\beta^S_{AA}\,\gamma^S_{AA}\,\alpha^2 + \beta^R_{aa}\,\gamma^R_{aa}\,\beta^S_{aa}\,\gamma^S_{aa}\,(1-2\,\alpha)}{\alpha\,(\beta^S_{AA}\,\gamma^S_{AA} - \beta^S_{aa}\,\gamma^S_{aa})} < \mu$
5. Disparition des AA, Aa et aaR → Fixation de a	N'existe jamais si $\beta^R_{aa}\,(1-\alpha)\,\gamma^S_{aa}\ \neq 0$	-	-
6. Disparition des AA, Aa et aaS → Fixation de a	N'existe jamais si $\beta^S_{aa}\,(1-\alpha)\,\gamma^R_{aa}\ \neq 0$	-	-
7. Disparition des AA et Aa → Fixation de a	Inconnue	Voir polynôme (39)	Jacobienne non étudiée
8. Disparition des AA, aaS et aaR	$\mu < \beta^S_{Aa}\,\gamma^S_{Aa}\,\alpha$	$E^4 = (0, \kappa^S\,(1 - \frac{\mu}{\alpha\,\beta^S_{Aa}\,\gamma^S_{Aa}}), 0, 0)$	Jamais stable
9. Disparition des AA et aaS	N'existe jamais si $mu \neq 0$	-	-
10. Disparition des AA et aaR	N'existe jamais si $\beta^S_{Aa}\,\gamma^S_{Aa} > \beta^S_{aa}\,\gamma^S_{aa}$	-	-
11. Disparition des Aa et aaS	N'existe jamais si $mu \neq 0$	-	-
12. Disparition des Aa et aaR	N'existe jamais si $\beta^S_{AA}\,\gamma^S_{AA} > \beta^S_{aa}\,\gamma^S_{aa}$	-	-
13. Disparition des aaS et aaR	N'existe jamais si$\beta^S_{AA}\,\gamma^S_{AA} > \beta^S_{Aa}\,\gamma^S_{Aa}$	-	-
14. Disparition des aaS	N'existe jamais si $\beta^S_{AA}\,\gamma^S_{AA} > \beta^S_{Aa}\,\gamma^S_{Aa}$	-	-
15. Disparition des aaR	N'existe jamais si $\beta^S_{AA}\,\gamma^S_{AA} > \beta^S_{Aa}\,\gamma^S_{Aa}$	-	-
16. Survie de l'ensemble des génotypes	N'existe jamais si $\beta^S_{AA}\,\gamma^S_{AA} > \beta^S_{Aa}\,\gamma^S_{Aa}$	-	-

2.2.4 Conclusion

Cette analyse de la phase épidémique clonale sur l'hôte télien nous permet de mettre en évidence l'existence de plusieurs équilibres possibles. L'existence et la stabilité de ces équilibres peuvent dépendre du taux de mortalité de la population d'agents pathogènes. On distingue deux taux de mortalité $0 < \mu_1 < \mu_2 < 1$. Un taux de mortalité supérieur à μ_2 mène à une extinction de population. Un taux de mortalité intermédiaire, *i.e.* entre μ_1 et μ_2, mène à une fixation de l'allèle d'avirulence A dans la population. Pour ces deux équilibres, aucun contournement de résistance n'a lieu et le compartiment d'hôtes résistants R reste inoccupé. De plus, l'équilibre de fixation de l'allèle A dans la population n'est stable que si les individus sur S ont un avantage sélectif par rapport aux individus sur R.

La fixation de l'allèle virulent a dans la population mène nécessairement ici à une coexistence des individus aa sur S et sur R, *via* les mouvements migratoires entre compartiments. Cela implique un contournement de résistance et une invasion du compartiment R par la population d'agents pathogènes.

Aucun équilibre de ce système ne mène à la survie d'individus hétérozygotes Aa. Cela est dû notamment aux conditions posées sur les différents avantages sélectifs des individus selon leur génotype (succès d'implantation des spores et taux de sporulation) : les individus hétérozygotes ont toujours une valeur sélective plus faible que les individus homozygotes AA. Ils disparaissent donc au profit des homozygotes avirulents, et n'ont pas la possibilité (contrairement aux homozygotes virulents aa) de survivre en contrebalançant cette faible valeur sélective par l'invasion d'un nouveau compartiment.

L'étude de la phase discrète du modèle correspondant à la reproduction sexuée sur un hôte alternant n'est pas présentée car nous n'avons pas réussi à résoudre les systèmes correspondants. Cette phase temporelle discrète permettrait cependant d'expliquer l'existence d'une population d'hétérozygotes Aa au début de chaque année. En effet, si des individus AA et aa coexistent à la fin d'une saison épidémique, nous nous attendons à ce que le brassage allélique lors de la reproduction sexuée entraîne la ré-émergence de génotypes hétérozygotes Aa.

2.3 Conclusion en regard du contournement de la résistance RMlp7 par *Melampsora larici-populina*

Nous avons montré dans la première partie de ce chapitre que la ploïdie et le cycle de vie de l'agent pathogène jouent un rôle majeur dans la probabilité d'aboutir à un contournement de résistance et dans la durabilité de cette résistance. L'agent pathogène Mlp est diploïde, et accomplit son cycle de vie avec une alternance d'hôtes chaque année. D'après notre modèle de simulation, ces deux conditions sont celles qui aboutissent aux trajectoires

évolutives et démographiques les plus stochastiques et donc les plus imprévisibles. La stochasticité observée chez les agents pathogènes diploïdes est due à l'état particulièrement vulnérable de l'allèle virulent lorsqu'il se trouve à l'état hétérozygote. Au sein d'un individu phénotypiquement avirulent, l'allèle virulent est uniquement soumis à la dérive génétique, ce qui peut entraîner des trajectoires particulièrement variables. Cet effet est accentué pour les faibles tailles de populations. L'ajout d'un coût sélectif à l'allèle virulent dans la modélisation analytique entraîne une nouvelle source de vulnérabilité de cet état hétérozygote. Avec ce coût sélectif, les individus hétérozygotes sont désavantagés quel que soit l'hôte sur lequel ils se trouvent. Par conséquent, nous n'observons aucun équilibre stable correspondant au maintien d'individus hétérozygotes dans la population lors des cycles de reproduction clonale. Ces individus hétérozygotes peuvent ré-émerger chaque année lors du brassage allélique de la reproduction sexuée. Ce phénomène est accentué pour un cycle de vie avec alternance d'hôtes, pour lequel la reproduction sexuée réunit sur l'hôte écidien l'ensemble des individus se reproduisant sur hôtes sensibles et résistants.

La modélisation par simulations met en évidence l'importance de la proportion d'hôtes résistants dans le paysage sur la durabilité d'une résistance. Une forte proportion d'hôtes résistants diminue la probabilité de contournement, mais augmente en même temps la vitesse d'expansion de la virulence dans les rares cas où le contournement a lieu. Suite au déploiement massif du cultivar Beaupré, la proportion d'hôtes résistants commercialisés pouvait atteindre 80% localement dans les années précédant le contournement de la résistance RMlp7. Ce paramètre peut ainsi expliquer l'expansion fulgurante de la virulence 7 sur l'ensemble du territoire français entre 1994 et 1998, à partir de l'invasion des premiers hôtes résistants au nord de la France.

La fréquence initiale d'allèles virulents dans la population augmente la probabilité d'aboutir à un contournement de résistance. La caractérisation du locus candidat AvrMlp7 montre que les deux allèles virulents avr et Ø étaient présents tous les deux à l'état hétérozygote dans les premières populations de Mlp échantillonnées en 1990, soit 4 années avant le contournement de résistance (LOUET *et al.*, 2021). L'existence de deux mécanismes d'altération distincts et compatibles chez Mlp revient à doubler la fréquence initiale d'allèles virulents dans la population, ce qui a pu augmenter la probabilité d'aboutir à un contournement de la résistance RMlp7.

Ces deux modèles mettent en évidence le lien étroit entre les trajectoires démographiques et évolutives d'une population lors d'un tel évènement de sélection. Le contournement de résistance donne accès à un nouveau territoire et entraîne une expansion démographique qui peut être considérable si la proportion d'hôtes résistants est élevée. Dans les cas les plus extrêmes, l'adaptation de l'agent pathogène à l'hôte résistant peut contrebalancer un déclin démographique de la population qui aurait autrement mené à une extinction. Ces phénomènes de sauvetage évolutif illustrent une interdépendance particulièrement

forte entre les trajectoires démographiques et évolutives. Ce phénomène est majoritairement observé pour des agents pathogènes diploïdes avec alternance d'hôtes, aux faibles taux de croissance et pour de fortes proportions d'hôtes résistants. Ce scénario pourrait correspondre à notre cas d'étude, sous réserve d'un taux de croissance de Mlp suffisamment bas. Malgré la forte proportion de cultivar Beaupré dans le paysage populicole, il paraît très improbable que la population de Mlp ait été sur le point de s'éteindre avant le contournement de la résistance RMlp7. La survie de Mlp a pu être favorisée par un taux de croissance élevé et par les régions où l'hôte écidien et les peupliers sensibles coexistent sans plantations proches de peupliers résistants. Au delà du système cultivé, l'hôte sauvage (le peuplier noir *Populus nigra*) qui prédomine dans certaines vallées de l'arc Alpin et proches du Mélézin offre un refuge certain aux populations de Mlp. Une manière de dissocier les différents scénarios démographiques lors d'un contournement de résistance pourrait être de s'intéresser à l'évolution des structures génétiques qui en résultent.

Chapitre 3

Etude démogénétique d'un contournement de résistance

Le chapitre précédent nous a permis d'expliquer l'expansion particulièrement rapide de la virulence 7 dans les populations de Mlp au moment du contournement de la résistance RMlp7. Ce contournement s'est accompagné du remplacement du groupe génétique « Fossile » par un nouveau groupe génétique « Cultivé » dans les populations de Mlp sur l'ensemble du territoire français. Une telle évolution de la structure génétique neutre de Mlp au cours de ce contournement reste encore inexpliquée.

L'objectif de ce chapitre est de caractériser les différentes signatures génétiques neutres pouvant émerger dans une population d'agents pathogènes lors d'un contournement de résistance et de déterminer leurs conditions d'observation. Je m'intéresse en particulier à la diversité de ces signatures génétiques et à leur lien avec les variations de tailles de populations induites par un contournement de résistance.

3.1 Article 3. Interplay between demography and selection: a forward model illuminates the temporal genetic signatures of rapid adaptation in plant pathogens

Pour caractériser l'évolution génétique d'une population d'agents pathogènes au cours d'un contournement de résistance, je reprends dans cette partie le modèle par simulations développé dans le Chapitre 2.1, en y ajoutant l'évolution de marqueurs génétiques neutres. En plus du locus d'avirulence sous sélection, je modélise ainsi pour chaque individu 100 loci neutres caractérisés par 4 états alléliques possibles. Ces marqueurs génétiques neutres permettent de calculer un ensemble d'indices génétiques intra et inter-populationnels et de suivre leur évolution au cours du temps, conjointement à l'évènement de contournement.

Ces résultats sont ici présentés sous la forme d'un article de recherche en préparation intitulé « Interplay between demography and selection: a forward model illuminates the temporal genetic signatures of rapid adaptation in plant pathogens » et qui sera soumis à la revue HEREDITY prochainement.

Interplay between demography and selection: a forward model illuminates the temporal genetic signatures of rapid adaptation in plant pathogens

Méline Saubin[1], Solenn Stoeckel[2], Aurélien Tellier[3] and Fabien Halkett[1]

[1] Université de Lorraine, INRAE, IAM, F-54000 Nancy, France

[2] INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, F-35653 Le Rheu, France

[3] Professorship for Population Genetics, Technical University of Munich, Freising, Germany

1 Introduction

Disease outbreaks caused by pathogens impact both natural and human-managed ecosystems (*e.g.* agro-ecosystems) (Anderson et al., 2004; Tobin, 2015; Savary et al., 2019). The number of emerging diseases is increasing exponentially and unprecedentedly during the last decades (Fisher et al., 2012). Understanding pathogen evolution is essential to comprehend how they affect ecosystems (Fischhoff et al., 2020) and to establish relevant disease management programs (Bonneaud and Longdon, 2020). Yet, this task is particularly challenging due to the rapid adaptation of pathogen populations (McDonald and Linde, 2002; Saubin et al., 2022), and the high stochasticity of pathogen evolutionary trajectories (Parsons et al., 2018). In addition, pathogens depend on host abundance over time and space (Stukenbrock and McDonald, 2008) with intricate life cycles and reproductive modes (Agrios, 2005) that combine to shape distinctive patterns of genetic structure (*e.g.* Rouger et al., 2016).

Pathogen populations commonly endure large demographic variations, including repeated bottlenecks and founder events. Most pathogen species display seasonal growing patterns with marked differences between epidemic and dormant phases (Agrios, 2005). The latter phase corresponds to a surviving period to face harsh environmental conditions or host absence, during which the pathogen population commonly collapses. These extreme variations in population sizes make pathogen populations particularly vulnerable to local extinctions (Laine and Barrès, 2013). Moreover, pathogen populations are subjected to strong selection, in particular during host adaptation events (Gladieux et al., 2011). Selection affects the pathogen's demography, increasing mortality of the non-adapted part of pathogen populations and leading to a burst of adapted pathogens (McDonald and Linde, 2002). Selection pressures are all the more strengthened in human-managed ecosystems such as in agrosystems (Stukenbrock and McDonald, 2008; Zhan et al., 2015).

In some extreme cases, mechanisms can lead to the pathogen survival even if the decline of the population was approaching extinction. Three forms of such population 'rescue' are commonly described (Carlson et al., 2014): 1) the demographic rescue, when the population survival is only attributed to the increase in population size due to immigration of new individuals (Brown and Kodric-Brown, 1977); 2) the genetic rescue, when the survival of the population is attributed to the novel genetic variation brought by the immigration of new individuals, in a small population suffering genetic load (Thrall, 1998); 3) the evolutionary rescue, when adaptive evolution lead to population recovery from negative growth initiated by environmental change (Gonzalez et al., 2013; Bell, 2017). These forms of rescue closely link demography and selection. The probability of fixation of alleles is strongly impacted by changes in population size (Otto and Whitlock, 1997), with the effect of genetic drift accentuated by a reduction in population size. It is therefore all the more important to focus on the interplay between selection and genetic drift in a population with fluctuating size (Gokhale et al., 2013; Živković et al., 2019) to weigh up the balance between deterministic and stochastic processes that drive the evolutionary trajectories of pathogen populations.

The joint study of demographic variations and genetic evolution of the pathogen population can be achieved using population genetics (Day and Proulx, 2004; Day and Gandon, 2007; Saubin et al., 2022). Coupling epidemiology and population genetics provides insights into the mechanisms underpinning pathogen evolution or host-parasite coevolution and acting at both short (ecological) and long (evolutionary) time scales (Milgroom and Peever, 2003; Archie et al., 2009). This framework is thus dedicated to the study of genetic signatures of strong and rapid selective events (Saubin et al., 2022). The interplay between demography and selection is captured by monitoring both selected and neutral loci. It allows in particular detailed analyses of transition periods (Day and Proulx, 2004; Day and Gandon, 2007; Bolker et al., 2010), through variables like the pathogen population size, affecting both the disease incidence in epidemiology and the impact of genetic drift in population genetics (McDonald, 2004; Živković et al., 2019).

Pathogen adaptation to plant genetic resistances is a typical example of rapid adaptation linked to variations in population size. In agrosystems, the breeding of resistant hosts based on the gene-for-gene interaction is crucial to mitigate the severity of epidemics of plant pathogens. Gene-for-gene interaction attracts a great deal of breeding efforts because in most cases it confers complete host immunity. In the simplest form, plant genetic resistance thus prevents infection from a class of pathogen genotypes called *avirulent* (or non-infective). In agrosystems, the deployment of pure resistant plants exerts a strong selection pressure on pathogen populations that favours any variant that can infect the resistant host (Zhan et al., 2015). This class of pathogen genotypes is called *virulent* (or infective). The infective phenotype is usually determined by changes at a single locus, the avirulence locus. Therefore, the issue of qualitative resistance deployments is often a resistance overcoming, *i.e.* the invasion of resistant hosts by virulent individuals (Johnson, 1984; Pink and Puddephat, 1999; Brown and Tellier, 2011; Burdon et al., 2016). The resistant variety is usually deployed widely across the landscape, and this intense and unidirectional selection pressure can result in severe epidemics, which impose rapid and drastic demographic changes for the pathogen population (Burdon et al., 2016; Persoons et al., 2017;

Saubin et al., 2021). We have limited knowledge on how such events shape the genetic evolution of pathogens and what would be the typical genetic signatures of such strong demographic and selective events on the pathogen population. This theoretical knowledge would help to decipher the causes of current and past events of resistance overcoming and, ultimately, to predict when and how often such events occur.

In this study, we aim to understand the interplay between demography and genetic evolution at the avirulence locus and neutral markers during resistance overcoming, and to assess the impact of such rapid adaptive events on the genetic structure of the pathogen population. We model pathogen population dynamics and genetic evolution to investigate the impact of the pathogen life cycle on these selective and demographic dynamics using a demogenetic approach. We perform simulations under several realistic scenarios of resistance overcoming. Using a clustering method dedicated to time-series variations, we show that we are able to identify several main scenarios of eco-evolutionary dynamics. We conclude by commenting on the typology of these dynamics and the potential to use our simulation framework to analyse real datasets.

2 Materials and methods

2.1 Model description

2.1.1 Model overview

To study the effects of resistance overcoming on genetic indices in diploid pathogen populations along generations, we develop an individual-based, forward-time demogenetic model. It couples population dynamics and population genetics to follow through time the evolutionary trajectory of different genotypes at the selected avirulence locus and at neutral genetic markers scattered in the genome. The model is similar to the model described in Saubin et al. (2021), with the addition in this study of the evolution of neutral markers through time. We consider a life cycle commonly found in temperate pathogen species, which alternate rounds of clonal reproduction with an annual event of sexual reproduction (Agrios, 2005). This model is designed to simulate two distinct life cycles: with or without host alternation for the sexual reproduction. Without host alternation, the model simulates the evolution over time of a population of pathogens on two static host compartments of the same host species: a resistant compartment (R) and a susceptible compartment (S). Fixed carrying capacities of pathogens, K_R and K_S, are respectively assigned to R and S compartments and represent the maximum amount of pathogens that each compartment can sustain. With host alternation, another alternate host species (a third compartment, A) is added to the simulations. Pathogens annually gather on this additional compartment to achieve sexual reproduction. This static compartment is assumed to be larger than the two other compartments, with a fixed carrying capacity K_A. Hereafter, we refer to the three compartments as S (susceptible hosts), R (resistant hosts) and A (alternate hosts). We do not consider spatial substructure among host compartments. The way we model the life cycles is presented in Figure 1, and can be described as follow.

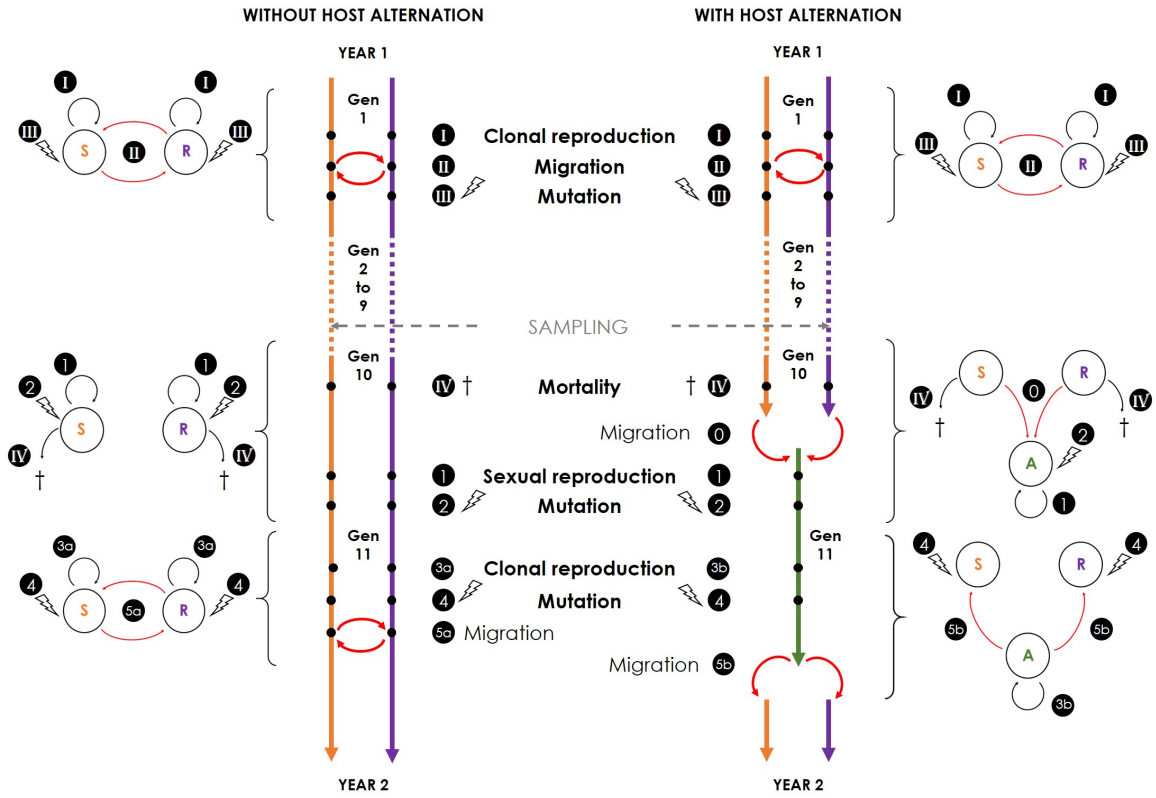


Figure 1: Modelling steps for each simulated year with the three host compartments S, R, and A. Each year is composed of 11 generations. During the clonal phase (generation 1 to 9), each generation is composed of three steps identical between both life cycles: (I) clonal reproduction (equation 2); (II) migration of a proportion m of each population between R and S; (III) mutation at all neutral markers with a mutation rate of 10^{-3}. The sampling takes place at the end of generation 9. At the end of the clonal phase, the pathogen overwinter as a dormant stage and is subjected to (IV) mortality of a proportion τ of each population. Then the sexual phase (generations 10 and 11) differs depending on the life cycle: (0) represents the migration of all individuals from R and S towards A; (1) sexual reproduction (equation 1); (2) and (4) mutation of all neutral markers with a mutation rate of 10^{-3}; (3a) and (3b) clonal reproduction (equation 2); (5a) migration of a proportion m of each population between R and S; (5b) migration of all individuals from A towards R and S.

2.1.2 Reproduction events

Each discrete generation corresponds to a reproduction event, either sexual or clonal. Each year is composed of g non-overlapping generations, with one annual sexual reproduction event followed by a succession of $g - 1$ clonal reproduction events. In our simulations, we consider a year composed of $g = 11$ generations, based on the case for the poplar rust pathogen with host alternation (*Melampsora larici-populina*, Hacquard et al., 2011) and the apple scab pathogen without host alternation (*Venturia inaequalis*, Le Cam, personal communication). At each reproduction event, successive generations do not overlap so that parents give way to offsprings and the new population is composed exclusively of the newly produced individuals.

Before each sexual reproduction event, a proportion τ of pathogen individuals is picked randomly to form the parental population. We fixed $\tau = 0.75$ to mimic pathogen life cycles displaying a moderate reduction in population size before and during sexual reproduction that usually takes place before or

during the unfavourable season, such as winter in temperate climates.

For the sexual reproduction event itself, the population size is considered constant before and after the reproduction event:

$$N_{n+1} = N_n \tag{1}$$

With N_{n+1} the population size at generation $n+1$ and N_n the population size at generation n. Sexual offspring genotypes are obtained by random association of gametes produced by the parental population.

For each clonal reproduction event on S, R, or A, the population growth has a logistic function, with the following relation:

$$N_{n+1} = N_n + (r-1) \times N_n \times (1 - \frac{N_n}{K}) \tag{2}$$

With r the growth rate of the pathogen population, K the carrying capacity of the compartment (K_S, K_R, or K_A for S, R, or A respectively). For all clonal reproduction events, offspring genotypes are drawn randomly with replacement from the parental population in which all parental pathogen individuals have the same uniform probability to reproduce.

2.1.3 Migration events

A regular two-way migration event takes place each generation before the reproduction event, between R and S. The number of migrants is determined by a fixed migration rate (m) multiplied by the number of individuals on the source compartment. In this effective migration scheme, migrants are always successfully established on the destination compartment, even if the number of individuals on this compartment reached its maximum carrying capacity. Thereby, this choice enables the immigration of new pathogens regardless of the size of the population, as it is observed in natural populations for plant pathogens.

For the life cycle with host alternation, the annual sexual reproduction event coincides with the obligate migration of the entire pathogen population to and from the alternate host. The first migration event takes place once per year after $g-2$ clonal reproduction events on R and S. For this migration event, an established proportion of individuals τ is picked randomly from R and S and migrates to A. R and S compartments are then emptied of any pathogen for the next two generations. After the two reproduction events (sexual and clonal) on A, the second migration event redistributes randomly all individuals from A to R and S, in proportion to the compartment carrying capacities (K_S and K_R), leaving no living individuals on A.

2.1.4 Genetic evolution of the pathogen population

To simulate the gene-for-gene interaction between resistant and susceptible hosts and virulent and avirulent pathogens, we model allele frequency evolution at the avirulence locus assuming standing genetic variation (Barrett and Schluter, 2008). In diploids, pathogens are virulent only if the products

of both alleles avoid detection by the host (Stukenbrock and McDonald, 2009; Thrall et al., 2016). In the following, we name Avr the dominant allele conferring avirulence and avr the recessive allele conferring virulence at the homozygous state.

In addition to the selected avirulence locus, we simulate the evolution of 100 neutral genetic markers with a mutation rate of 10^{-3}. Each locus has four possible allelic states under a classical k-allele mutation model (Wright, 1949). Upon mutation, an allele changes into any of the three other allelic states with equal probability. At these 100 loci, we compute 19 classical population genetics indices each year before the sexual reproduction on both R and S compartments (Table 1).

These indices are chosen to 1) describe intra-population genetic and genotypic diversity, 2) measure overall linkage disequilibrium, and 3) assess genetic differentiation (F_{ST}). The differentiation is considered between populations on R and S at a given generation, and through time between a population on R or S at a given generation and the initial genotypic state. The variation in these indices captures the footprints of the different processes expected to occur during resistance overcoming: effects of reproductive modes, founder events followed by expansions, patterns of coancestry and admixture between R and S.

2.1.5 Implementation of model analyses

The model is implemented in Python (version 3.7, van Rossum, 1995) and Numpy (Harris et al., 2020). Each simulation starts with genotypes randomly drawn from the four possible alleles followed by a burn-in period of 11,000 generations under a constant population size of K_S individuals. In this way, we ensure that the pathogen population is at the drift-mutation equilibrium before the resistance overcoming. A proportion f_{avr} of virulent alleles is introduced (replacing avirulent Avr alleles) on S after the burn-in period as initial standing genetic variation. Homozygous avr/avr and heterozygous AVR/avr individuals are therefore initially present, based on the frequency f_{avr}. Simulations are run with a fixed total carrying capacity for the host population sizes of each host species, $K = K_A = K_R + K_S =$ 10,000. The size of R is given by the fixed and constant proportion of resistant hosts $propR = \frac{K_R}{K}$ time K. The size of S is deduced from K and K_R.

We run a random design set of 30,000 simulations. Each combination of parameter values is drawn at random in defined distributions to investigate a wide range of genetic dynamics through resistance overcoming (Table 2). Simulations are run for 400 generations. During this period, nearly all replicates reach new steady states including invasion of R and loss of the avirulent Avr allele or extinction of the pathogen population. To focus on the genetic signatures of a resistance overcoming, only the simulations with at least 60% of virulent avr alleles at the end of the simulation are kept for this analysis. This threshold is chosen to ensure that the resistance overcoming and the invasion of the resistant compartment do occur during the simulated period.

To evaluate the effect of stochasticity on the dynamics, we perform also a regular simulation design composed of 100 replicates of 216 fixed combinations of parameter values (Table 2).

Table 1: Description of population genetics indices computed in the model and calculated each year before the sexual reproduction on both R and S compartments.

Index	**Description**	**Reference**	**Clustering**	**Graphics**
$\beta\ of Pareto$	Genotypic diversity index	Arnaud-Haond et al., 2007	X	
$Simpson\ index$	Genetic diversity index			
$Shannon - Wiener\ index$	Genetic diversity index			
G/N	Proportion of unique genotypes	Dorken and Eckert, 2001		
G	Number of unique genotypes	Halkett et al., 2005		
$Mean\ H_E$	Mean expected heterozygosity overall loci	Nei, 1978		X
$Variance\ H_E$	Variance of the expected heterozygosity overall loci			X
$Mean\ H_O$	Mean observed heterozygosity overall loci			
$Variance\ H_O$	Variance of the observed heterozygosity overall loci			
$Mean\ LA$	Mean number of alleles by locus	Nielsen and Signorovitch, 2003		X
$VarianceLA$	Variance of the number of alleles by locus			
$Mean\ F_{IS}$	Mean inbreeding coefficient overall loci	Wright, 1931, 1949	X	X
$Variance\ F_{IS}$	Variance of the inbreeding coefficient overall loci	Stoeckel and Masson, 2014	X	
$\bar{r}_D$	Linkage disequilibrium index	Agapow and Burt, 2001	X	X
N_e	Effective population size	Wright, 1931		
F_k	Pollack's standardised variance of allele frequency	Pollak, 1983	X	
F_c	Nei and Tajima's standardised variance of allele frequency	Nei and Tajima, 1981		
$F_{ST}\ R-S$	Population differentiation between populations on R and S compartments	Wright, 1949, 1978		X
$Temporal\ F_{ST}$	Population differentiation between the initial population and the population at each recorded generation		X	X

Table 2: Input parameters and their range of variations for the two simulation designs run for 400 generations. For the regular simulation design, 100 replicates are run for each combination of parameters.

Variable	Description	Random design		Regular design
		Distribution	Interval	
f_{avr}	Initial frequency of the avr allele in the pathogen's population	Log-uniform	[log(0.0005), log(0.3)]	**4 levels**: 0.01; 0.03; 0.1; 0.32
m	Migration rate between R and S compartments	Log-uniform	[log(0.001), log(0.2)]	**3 levels**: 0.001; 0.01; 0.1
r	Growth rate of the pathogen	Uniform	[1.1, 2]	**3 levels** : 1.2; 1.6; 2.0
$propR$	Proportion of resistant hosts in the landscape	Uniform	]0.01, 0.99[	**3 levels**: 0.1; 0.5; 0.9
$Cycle$	Life cycle of the pathogen	Binomial	Without or with host alternation, probability 0.5	**2 levels**: Without host alternation; With host alternation

2.2 Comparisons of temporal dynamics

Analyses of changes in population genetics indices are performed using R (R, 2018). We present all results on S because it enables to compare the effect of life cycles, all else being equal, and to represent the genetic signatures expected without selection. When needed, we refer to the evolution of indices on R in supplementary data.

To analyse the dynamics of population genetics indices from the random simulation design, we performed clustering analyses regrouping simulations leading to similar dynamics on S. Distinct clustering analyses are performed for the two life cycles using the R package DTWCLUST (Sardá-Espinosa, 2019). To avoid redundancy in genetic information, clustering analyses are based on the temporal dynamics of the six indices that are orthogonal by their mathematical construction: $\beta\, of Pareto$ for genotypic diversity, $\bar{r}_D$ for overall linkage disequilibrium between loci, $Mean F_{IS}$ and $Variance F_{IS}$ for allele identity within individuals, F_k for the variation of allele identity between individuals of a same compartment, and $Temporal\ F_{ST}$ for the genetic differentiation. We perform a multivariate analysis by merging the normalised temporal dynamics of the six indices. We build a distance matrix between all simulations based on the dynamic time warping distance ('DTW_basic'), with the norm 'Euclidean distance' for the local cost matrix and the step pattern 'symmetric2'. We perform hierarchical clustering based on the distance matrix, with the agglomeration method 'Ward.D2'. We compare hierarchical clustering for a number of clusters between 2 and 8 included. For each life cycle, we select the number of clusters that maximises the Calinski-Harabasz index (Arbelaitz et al., 2013).

To understand how input parameters impact the clustering, we represent the distribution of input parameters for each cluster and both life cycles. We assess significant differences between distributions of parameter values using pairwise Kruskal-Wallis tests. To rank the impact of input parameters on the clustering, we perform a dominance analysis for each clustering and represent the estimated general dominance of each parameter on cluster assignments. Dominance analyses for both life cycles are performed with the R package DOMIR (Luchman, 2022).

For each cluster, we illustrate the evolution of population genetics indices through time. For the sake of clarity and to limit the amount of information, we focus in this article on the most informative four intra-population and two inter-population indices: $\bar{r}_D$, $Mean\ F_{IS}$, $Mean\ H_E$, $Mean\ LA$, $Temporal\ F_{ST}$, $F_{ST}\ R-S$ (Table 1). We display for each cluster and both life cycles the mean and standard error of these six population genetics indices calculated at each generation. These are not specific dynamics but mean changes over all simulations assigned to a given cluster.

To illustrate the realised dynamics of each population genetics index, we complement these results by displaying a representative realised simulation of each cluster (the medoid, *i.e.* the simulation that minimises the average distance to all other simulations in the same cluster). To highlight the effect of selection, we supplement each representation of the medoid dynamics with the corresponding null model dynamics, that is a simulation run under the same set of parameter values but without selection ($f_{avr} = 0$). The deviation between medoid and null model dynamics highlights the specific signatures of selection. To interpret genetic changes with respect to resistance overcoming, we calculate the generation of resistance overcoming as the generation for which 1% of R is invaded by virulent

individuals for the first time in a simulation. In addition, we define the generation of invasion as the first generation in which a virulent individual migrates to R.

Some simulations lead to evolutionary rescue, that is the invasion of R resulting in the recovery of the population collapse on S and preventing population extinction. To understand which input parameter combinations favour the occurrence of evolutionary rescue, we calculate for each simulation the growth rate threshold under which the population goes extinct if R is not accessible (population extinction in the 'null model'). We thus obtain the proportions of evolutionary rescue events in each cluster and perform a Fisher's exact test to assess the significance of the assignments to different clusters.

For the regular simulation design, we perform similar clustering analyses with and without host alternation to identify the proportion of simulations that result from the same combination of input parameters (replicates) but are assigned to different clusters.

3 Results

3.1 Influence of model parameters on the evolutionary dynamics

The clustering based on the random simulation design results in a partition into two clusters without host alternation and into three clusters with host alternation. For both life cycles, we name Cluster 1 the cluster that regroups the majority of the simulations (72 % and 78 % of the simulations without and with host alternation, respectively, Table 3). Cluster 1 displays small genetic changes through time for nearly all indices (Figure 2). Conversely, the other clusters display stronger changes in population genetics indices (Figure 2, Appendix Figures S1, S2).

Table 3: Cluster description for the clustering based on the random simulation design. This table contains the number of simulations assigned to each cluster, and the percentages of simulations leading to evolutionary rescue. Only simulations leading to resistance overcoming have been kept for the clustering analysis. The $P - value$ corresponds to the $P - value$ of the exact test of Fisher to determine if differences in proportions of evolutionary rescue events among clusters are significant.

Cycle	**Cluster**	**Nb of simulations**	**Evolutionary rescue (%)**	**Fisher** $P - value$
Without host alternation	1	4067	1.3	4.6×10^{-5}
	2	1575	3.0	
With host alternation	1	5858	3.7	$< 2.2 \times 10^{-16}$
	2	1581	10.6	
	3	81	97.5	

To understand the origin of the different types of dynamics, we jointly examine the influence of the input parameters on cluster delineation (Figure 3) and the difference in the distributions of parameter values for each cluster (Figure 4). Without host alternation, the most influential parameter for cluster delineation is the proportion of resistant hosts ($propR$), followed by the initial frequency of avirulent allele (f_{avr}), and, to a lesser extent, the migration rate (m) (Figure 3). As such, Cluster 2 is composed of simulations with significantly larger values of $propR$ and significantly smaller values of f_{avr} and slightly, albeit significantly, lower m. The growth rate (r) has no effect on cluster delineation (and consistently there is no significant difference in parameter distribution across clusters). It is important however to keep in mind that the analysis only considers simulations where R is successfully invaded

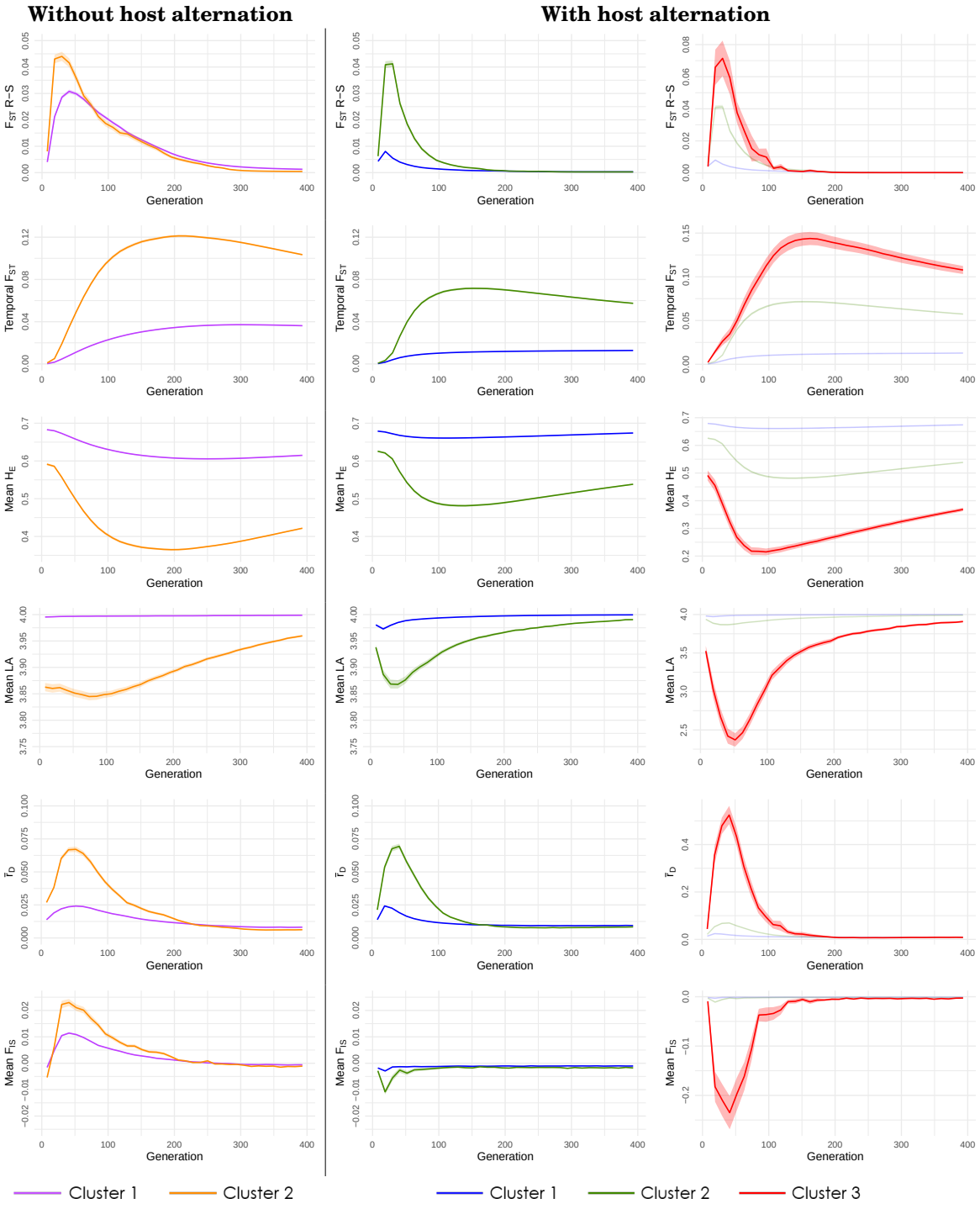
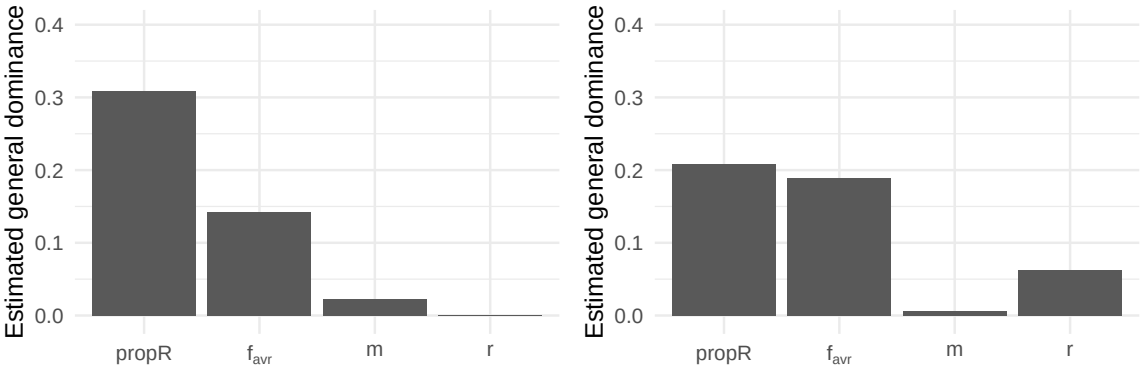


Figure 2: Temporal evolution of population genetics indices used for the clustering analysis. The plotted results correspond to the mean temporal dynamics for all simulations in the corresponding cluster, among all simulations of the random simulation design. Coloured intervals correspond to standard error intervals. For clarity, we apply the same scale for Cluster 1 and Cluster 2 of both life cycles, but we use a different scale to display the changes in population genetics indices for Cluster 3 with host alternation.

by the pathogen.



(a) Without host alternation

(b) With host alternation

Figure 3: Estimated general dominance calculated from a multinomial logistic regression applied to the cluster assignment for each simulation of the random design. The general dominance was estimated for the four input parameters: $propR$, f_{avr}, m, and r.

With host alternation, we observe the same differences between Cluster 1 and Cluster 2, with similar variations in the distribution of parameter values and ranking of parameter effects. The only difference is that r has an effect on cluster delineation (even more pronounced than m), with significantly higher growth rates for Cluster 2. Cluster 3 represents a particular case with skewed distributions towards very high values of $propR$ and low values of r (Figure 4). Cluster 3 displays rather high values of f_{avr} and higher values of m compared to the two other clusters.

3.2 Cluster delineation reflects different demographic scenarios

Cluster delineation, and thus the magnitude of genetic changes, reflects and distinguishes three demographic scenarios.

As seen above, the most influential parameter for cluster delineation is $propR$. As this parameter determines the maximal population sizes on R and S, it results in differences in population dynamics among clusters (Figure 5, Appendix Figure S3). The mean compartment size of S is higher for Cluster 1, irrespective of the life cycle, which leads to higher initial population sizes and less genetic changes through time on S.

For both life cycles, the second most influential parameter is the initial frequency of avirulent allele (f_{avr}), which plays a major role in the assignments to Cluster 2, with an over-representation of simulations with low values of f_{avr} (Figures 4, Appendix Figure S4). These low values of f_{avr} result in fewer virulent individuals. It causes founder effects on R (Figure 5) and leads to more pronounced genetic changes through time (Appendix, Figures S1, S2).

A peculiar range of parameter values defines Cluster 3 observed with host alternation. It results in evolutionary rescue dynamics. The low growth rate associated with a restricted size of S (high $propR$) causes an initial decrease in population size until near extinction (Figure 5). Then, the emergence of virulent individuals leads to the establishment of a new population on R and these individuals, by

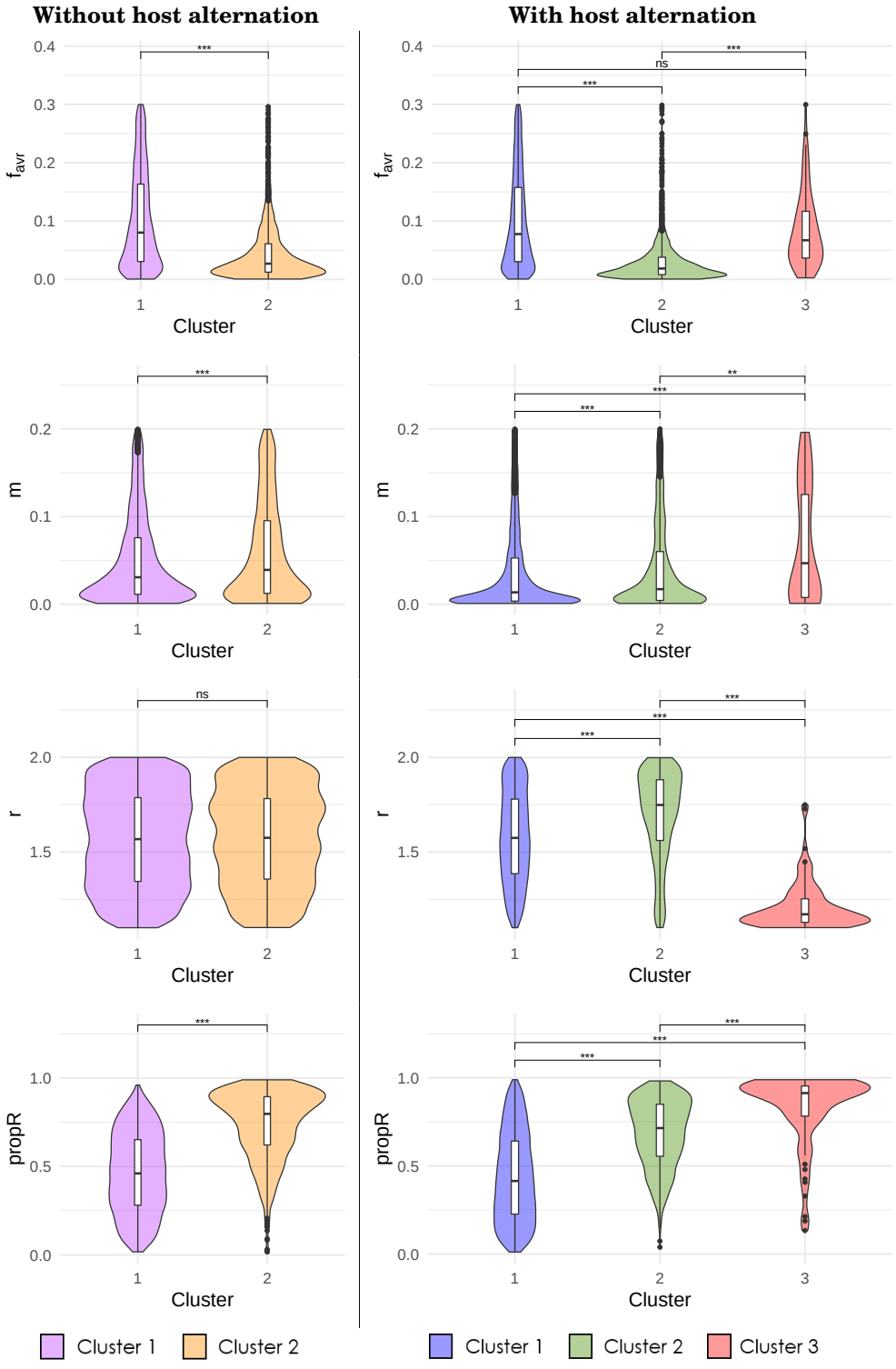


Figure 4: Distribution of epidemic parameters for each cluster obtained from the dynamics of population genetic indices. Difference between distributions were statistically assessed using pairwise Kruskal-Wallis tests : ns, non significant; ·, $P - value < 0.1$; *, $P - value < 0.05$; **, $P - value < 0.01$; ***, $P - value < 0.001$. Without host alternation, Cluster 1 and Cluster 2 are composed of 4067 and 1575 simulations, respectively. With host alternation, Cluster 1, Cluster 2 and Cluster 3 are composed of 5858, 1581, and 81 simulations, respectively.

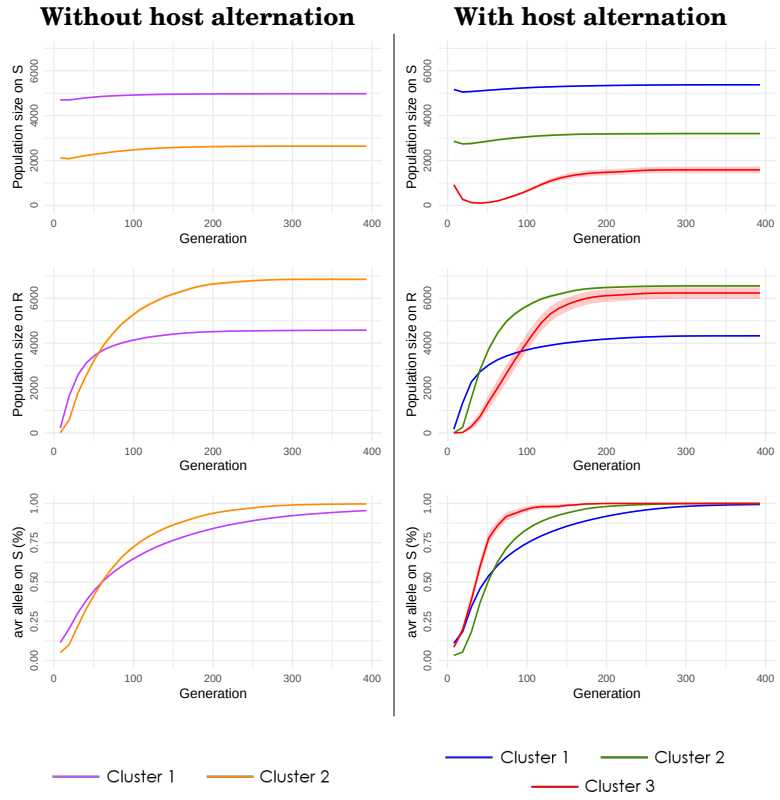


Figure 5: Temporal evolution of population sizes and virulent allele frequency depending on the cluster. The plotted results correspond to the mean temporal dynamics for all simulations in the corresponding cluster, among all simulations of the random simulation design. Coloured intervals correspond to standard error intervals.

migrating back to S, prevent population extinction (see example in Figure 6, Cluster 3). Proportions of evolutionary rescue events differ significantly among clusters, with Cluster 3 composed almost entirely of such simulations (Table 3).

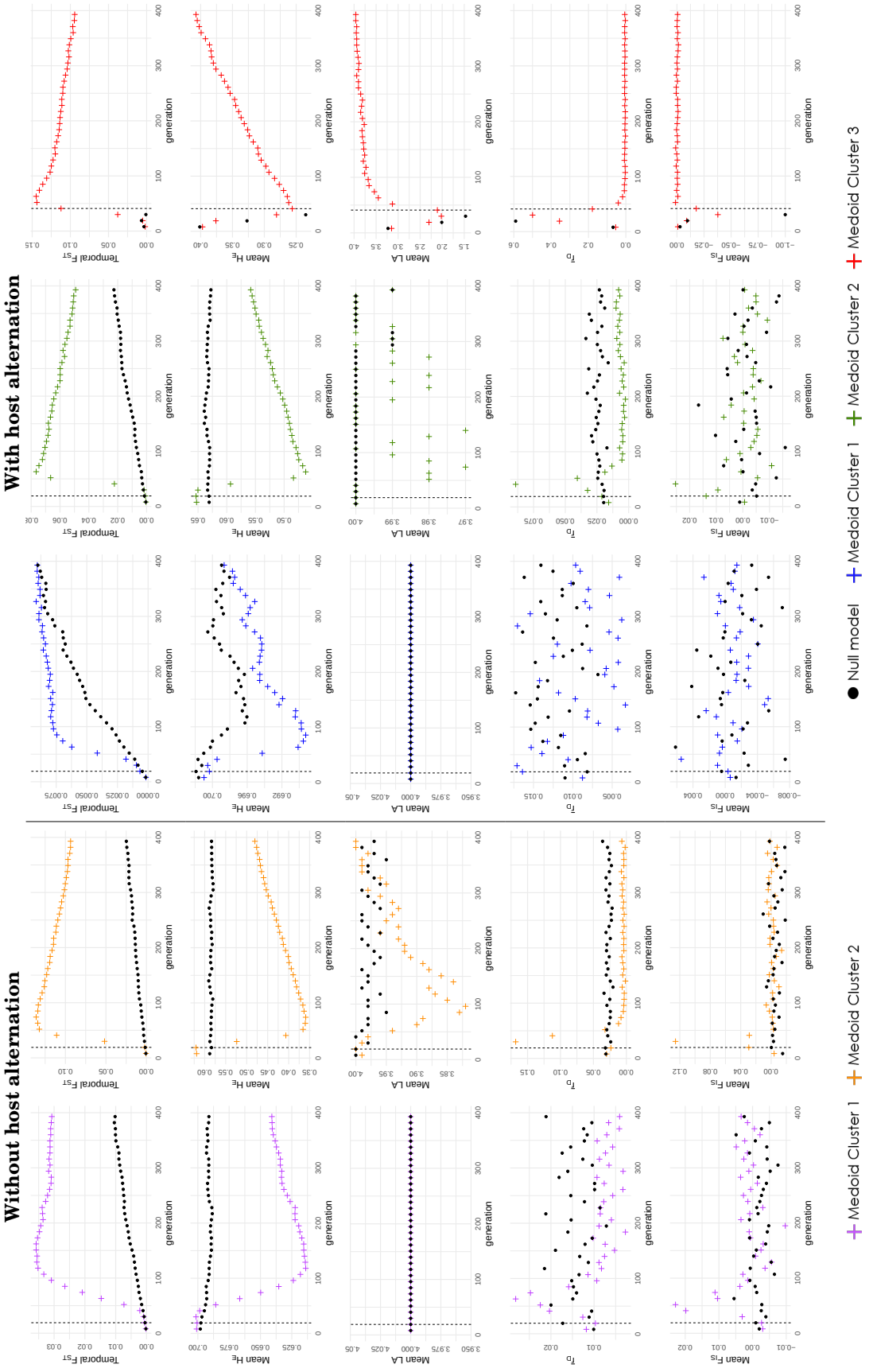


Figure 6: Comparison of the temporal evolution of population genetics indices used for the clustering analysis, between medoid simulations and null models. The plotted results correspond to the medoid dynamic of each cluster, with ('Medoid', cross shapes) or without ('Null model', round shapes) introducing virulent allele, that is, with or without resistance overcoming. For each of the plotted medoid, the vertical dashed line corresponds to the generation of resistance overcoming. Cluster 3 with host alternation displays no round shapes after the generation of resistance overcoming because the population goes extinct in the 'Null model'.

3.3 Inter- and intra-population genetic changes

To compare inter- and intra-population genetic changes associated with resistance overcoming, we focus on 1) mean dynamics, *i.e.* mean temporal evolution of population genetics indices on each cluster, calculated from neutral loci, and 2) realised dynamics, *i.e.* temporal evolution of population genetics indices for medoid or null model simulations of each cluster, calculated from neutral loci.

Inter-population genetic signatures

All clusters show a peak in the mean dynamics of genetic differentiation between populations on R and S ($F_{ST}R-S$, Figure 2). This peak in differentiation occurs within the first 50 generations and is concomitant with resistance overcoming (Appendix, Figure S5). The peak is more abrupt with host alternation, especially for Cluster 1, and the genetic differentiation persists for a shorter period with host alternation. This genetic differentiation decreases at the same rate for Cluster 1 and Cluster 2. The peak is higher for Cluster 2 than Cluster 1, which reflects stronger founder effects resulting from the invasion of R highlighted in Cluster 2. With host alternation, this peak in differentiation is magnified in Cluster 3 by the conjunction of the founder effect and the strong genetic drift on S resulting from the initial decrease in population size before evolutionary rescue.

For all clusters we observe a two-phase dynamics of temporal differentiation on S (from the initial population collected after burn-in) (Figures 2, 6). The first phase corresponds to an initial increase in temporal F_{ST} that coincides with resistance overcoming. Temporal F_{ST} reaches a maximum between 50 and 200 generations, depending on the cluster and the life cycle. This increase in temporal F_{ST} on S is delayed but similar to the temporal F_{ST} on R (see the example of the temporal F_{ST} on R for Cluster 2 with host alternation, Appendix, Figure S6). The second phase is more stable with a slight increase or decrease in temporal F_{ST}. This second phase is concomitant with the regain in genetic diversity on both compartments, which favours homogenisation of allele frequencies distorted by the founder event (Figure 2). With host alternation, simulations displaying strong genetic differentiation (realised dynamics of Cluster 2 and Cluster 3) show a steeper peak of differentiation (Figure 6). This is in accordance with the rapid decrease in differentiation between populations on R and S with host alternation, which shifts the temporal F_{ST} to the next phase of the dynamics. Conversely, without host alternation, differentiation between populations on R and S persists for a longer period, which leads to a slower change in temporal F_{ST} dynamics on S. Without selection, the temporal F_{ST} on S increases linearly through time at a slope depending on the relative importance of mutation and genetic drift forces (Realised dynamics of temporal F_{ST}, Figure 6). As the mutation rate is fixed, this slope depends on population size only. This increase without selection is much lower than that observed in the first phase of realised dynamics of resistance overcoming. The higher level of differentiation observed with selection is caused by genetic admixture with the differentiated population on R.

Intra-population genetic signatures

For all clusters, we observe a two-phase dynamics of temporal change in gene diversity (expected heterozygosity, H_E) on S. The first phase corresponds to a strong decrease in H_E, until a minimum value

is reached between 50 and 200 generations, followed by a slower increase towards a new mutation-drift equilibrium (Figure 2). However, the timing of this change differs among clusters. For Cluster 1 and Cluster 2 of both life cycles, the decrease in H_E follows the generation of resistance overcoming, while the null model dynamics show no variations in H_E (Figure 6). The decrease in gene diversity results from the immigration of less diverse pathogen individuals from the founded population on R that overcame the resistance. For Cluster 3 (with host alternation), a strong decrease in H_E is observed for both the null model and the medoid realised dynamics, and the drop in H_E precedes resistance overcoming. Unlike other clusters, this drop in H_E is preceded for Cluster 3 by a strong decrease in the mean number of alleles $Mean\ LA$ (Figures 2, 6), indicating a bottleneck.

Simulations display a peak in linkage disequilibrium, with a maximum value of $\bar{r}_D$ being reached in the first 100 generations (Figures 2, 6). The variation in $\bar{r}_D$ is to be examined in relation to the variation in F_{IS}, which patterns differ between clusters and life cycles. Without host alternation, $\bar{r}_D$ and F_{IS} display similar variations but slightly delayed in time, with a maximum value of F_{IS} reached in the first 50 generations. This indicates admixture of genetically differentiated individuals on S. With host alternation, F_{IS} is null or displays slightly negative values for Cluster 1 and Cluster 2 while $\bar{r}_D$ remains positive. This indicates that the signature of the admixture on S is rapidly being erased with host alternation. In Cluster 3 with host alternation, the peak in $\bar{r}_D$ comes with very negative values of F_{IS}. The negative values in F_{IS} results from the decrease in H_E that is preceding the decrease in H_O (Figure 6, Appendix, Figure S7).

3.4 Stochasticity of dynamics

As for the clustering based on the random simulation design above, temporal changes in population genetics indices from the regular simulation design group into two clusters without host alternation, and three clusters with host alternation (Table 4). For both life cycles, Cluster 1 regroups the majority of the simulations ($53\ \%$ and $83\ \%$ of the simulations without and with host alternation, respectively, Table 4). Notably, we identify simulation replicates (*i.e.* with identical combinations of parameter values) that show different outcomes in the temporal changes of population genetics indices. Hence, simulation replicates can be assigned to different clusters, demonstrating a certain degree of stochasticity. These replicates assigned to different clusters represent 6% and 12% of the simulations in Cluster 1 and Cluster 2 without host alternation, respectively, and 17%, 85% and 98% of the simulations in Cluster 1, Cluster 2, and Cluster 3 with host alternation, respectively (Table 4). Therefore, cluster delineation is not defined by specific combinations of parameter values, but rather by the inherent demographic scenarios underpinning the genetic dynamics.

Table 4: Cluster description for the clustering based on the regular simulation design. This table contains the number of simulations assigned to each cluster, and the percentages of simulations from a same set of parameters (100 replicates) which have been assigned to different clusters ('Different assignments (%)'). Only simulations leading to resistance overcoming have been kept for the clustering analysis.

Cycle	Cluster	Number of simulations	Different assignments (%)
Without host alternation	1	2966	5.9
	2	2605	12.4
With host alternation	1	6937	16.9
	2	1332	85.0
	3	42	97.6

4 Discussion

4.1 Genetic signatures of resistance overcoming

All the recorded population genetics indices were impacted by the simulated resistance overcoming. Overall, resistance overcoming leads to a founder effect on the resistant host, with a differentiated sub-sampled population settling and growing on resistant hosts. Migrations between susceptible and resistant hosts then homogenise the genetic coancestry over all pathogen populations at a pace that depends on the migration rate and pathogen population sizes. Resistance overcoming has a strong impact on the population genetic structure on susceptible hosts, with 1) a decrease in pathogen genetic diversity, 2) a peak in linkage disequilibrium, 3) a strong increase in temporal genetic differentiation between the initial and evolved populations, and 4) a peak in population differentiation between the susceptible and resistant compartments. The comparison of the evolution of the genetic indices through time with and without selection showed that these changes were signatures of evolution under selection and did not result from genetic drift only. Our first important result is thus that changes in population genetics of neutral markers allow identifying a selective event of resistance overcoming. We note that most of these genetic changes are transient, with a signature of resistance overcoming that vanishes in a few years only.

4.2 Typology of dynamics under resistance overcoming

Each evolutionary scenario represents distinct genetic signatures of resistance overcoming Simulations can be distinguished by the magnitude of their genetic signatures and grouped into clusters that are indicative of different evolutionary scenarios. This clustering is strongly linked to variations in pathogen population sizes. Cluster 1 regroups simulations with the slightest genetic signatures, associated with a steady and slow demographic expansion. Hence, a large part of the simulations leads to signatures of particularly low magnitudes. Cluster 2 regroups simulations with stronger genetic signatures, associated with larger demographic expansions on the resistant compartment. In extreme cases with host alternation, the few simulations assigned to Cluster 3 present the strongest genetic signatures, mainly associated with a specific demographic scenario, namely evolutionary rescue. These simulations are characterised not only by a strong demographic expansion on the resistant compartment (as for Cluster 2) but also by a demographic recovery on the susceptible compartment. Overall,

founder effects lead to a demographic expansion that is all the more important that the resistant host is abundant, because of the logistic population growth. During the first generations following the invasion of the resistant compartment, the successive clonal reproductions lead to a large population of few genotypes largely repeated, thus strongly entangling demographic variations and genetic changes.

Differences due to input parameters

In this model, the proportion of resistant hosts is the main determinant of pathogen population sizes, and not the intrinsic demographic parameters (pathogen growth rate and migration rate). For both life cycles, the proportion of resistant hosts determines the initial pathogen population size and the size of the compartment available through adaptation. As such, it drives the strength of both population expansion and selection pressure exerted on the pathogen population. With host alternation, this proportion also shapes the population demography in determining the vulnerability of the population to evolutionary rescue (See Figure 4 in Saubin et al., 2021). Pathogen control often leads to high proportions of resistant hosts in agricultural landscapes (Stukenbrock and McDonald, 2008; Zhan et al., 2015). Here we show that in case of resistance overcoming, higher proportions of resistant hosts lead to more pronounced genetic changes in the pathogen population. The initial proportion of virulent allele in the pathogen population is also a strong driver of the strength of the genetic signatures. This extent of standing genetic variation determines the proportion of individuals that will be able to respond to selection, in other words, the adaptive potential of the initial population. In particular, the proportion of virulent allele impacts the number of individuals that settle the virulent population, hence the genetic diversity of the founded population. A small proportion leads to fewer virulent individuals, and therefore a less diverse and more differentiated population founded on resistant hosts, as observed for frequent turnover of extinction and recolonisation (McCauley, 1991).

Differences due to the pathogen life cycle

Without host alternation, genetic signatures remain detectable for a longer period of time, because of sustained genetic admixture. Under this life cycle, gene flow results from the movements of a limited number of individuals, namely a proportion of the population determined by the migration rate. Migration occurs at each generation, which leads to a regular but progressive homogenisation of populations evolving on resistant and susceptible hosts. This accounts for the delay between the maximum differentiation observed between compartments at the time of the founder event, and the return to a low differentiation after the homogenisation of the populations. At the time of maximum differentiation, the immigration on susceptible hosts of individuals from the newly founded population on resistant hosts causes a Wahlund effect (Wahlund, 1928), that is a distortion of allele frequencies caused by the admixture of genotypes originating from different subpopulations. This explains the positive values of both the inbreeding coefficient and the linkage disequilibrium.

With host alternation, genetic differentiation between compartments fades more rapidly because of the obligate mating event taking place in a common host (A), where alleles at each locus are reshuffled

through sexual reproduction. It erases the Wahlund effect observed without host alternation as a single event of sexual reproduction among all individuals is sufficient for a return to Hardy-Weinberg equilibrium (Rouger et al., 2016). This explains the small or slightly negative values of the inbreeding coefficient. Yet, linkage disequilibrium remains positive because the associations of alleles across loci are still preserved for some time, as recombination occurs within individuals (whose allele frequencies are inherited from either population. Last, all individuals are redistributed randomly between the two compartments. The death of avirulent individuals on resistant hosts distorts allele frequencies and regenerates differentiation between pathogens evolving on resistant and susceptible hosts. This life cycle increases gene flow and leads to a fast homogenisation of pathogen populations.

These differences between life cycles explain the observed differences in demographic variations and genetic changes in pathogen populations. Examples of these different genetic outcomes can be found in the literature: from conservation of genetic structure without host alternation (*e.g.* in Leroy et al., 2013) to strong selective sweep and gene swamping with host alternation (*e.g.* in Persoons et al., 2017).

Stochasticity and genetic signatures

Beyond the influence of input parameters and life cycle on demographic variations and genetic changes, identical combinations of input parameters can lead to different outcomes. This variation among simulation replicates results from the inherent stochasticity of the model. In particular, migrants are picked randomly. The generation at which the first virulent individuals actually invade the resistant host drives the resulting demography and genetic structure. One year is composed of one sexual reproduction and ten clonal reproductions. Therefore, if the invasion occurs right after the sexual reproduction, several clonal reproductions follow, building a strongly differentiated population on resistant hosts. Conversely, if the invasion occurs at the end of the season of clonal reproductions, the population size does not change much before the next sexual reproduction reshuffles alleles in the population. Moreover, the stochasticity impacts not only the timing of invasion but also the number of successful migrations between comparments, hence the genetic diversity of the founded population. Beyond the inherent properties of the model, the life cycle with host alternation also adds stochasticity. Without host alternation, a first invasion of the resistant compartment settles a population, hence leading inevitably to resistance overcoming. Conversely, with host alternation, a first invasion event is necessary but not sufficient for a sustained settlement. The regular emptying of compartments before sexual reproduction and random individual redistribution can lead thereafter to stochastic invasions of resistant hosts as long as the virulent allele is rare (*e.g.* invaded compartment one year and cleared compartment the next year). As the virulent allele is recessive, it is more vulnerable to extinction with host alternation (Saubin et al., 2021). Here we show that in addition to the stochasticity in the fate of the virulent allele, this life cycle also increases the stochasticity in the genetic signatures of resistance overcoming.

4.3 Genetic signatures characteristic of evolutionary rescue

In extreme cases where the pathogen growth rate is not sufficient to counteract the mortality of individuals on resistant hosts, the population can decline towards extinction. Through resistance overcoming, a population rescue can occur. Such rescue can be considered as demographic or evolutionary, depending on the definition of the population. If we consider distinctly pathogen populations on susceptible and resistant hosts, the adaptation of virulent individuals leads to their invasion of the resistant host, hence to the survival of the population on the susceptible host. The survival of the population on the susceptible host corresponds to demographic rescue resulting from the immigration of adapted individuals from the resistant host. If we consider a single population gathering all individuals evolving on both hosts, the survival of the population corresponds to an evolutionary rescue event, *via* the adaptation to the new environment, *i.e.* the newly deployed resistant hosts. Such evolutionary rescue events occur mostly for the life cycle with host alternation (Saubin et al., 2021). This is because of additional mortality that originates once a year from the massive redistribution of individuals after the sexual reproduction, with the death of all avirulent individuals that migrate to resistant hosts. Besides the additional mortality, the massive redistribution also increases the probability that a virulent individual migrates to the resistant host, and hence increases the probability of evolutionary rescue (Saubin et al., 2021). Here we demonstrate that such events can lead to strong and typical genetic signatures at neutral loci, that define a specific cluster with host alternation (Cluster 3) and can thus be uncovered using population genetics indices. These dynamics are characterised by a bottleneck with few possible genotypes combining the remaining alleles. This is evidenced by the changes in several indices, such as the drop in the mean number of alleles, followed by the drastic reduction in gene diversity and the increase in linkage disequilibrium (Cornuet and Luikart, 1996). Unexpectedly, this bottleneck comes along with a strikingly negative value of inbreeding coefficient. To understand this result, we investigate the cause of the discrepancy between expected and observed heterozygosity. This is due to the clonal reproduction events that maintain identical genotypes, hence the value of observed heterozygosity remains constant over generations, whereas expected heterozygosity steadily decreases because the population size is collapsing very fast. We note that as the sampling takes place at the end of the clonal phase, the difference between expected and observed heterozygosity is magnified by the small genotypic drift that happened between clonal lineages over the handful of clonal generations. At the time of the founder effect on the resistant host, these dynamics are characterised by a large differentiation between compartments. Yet this differentiation erases as quickly as the other genetic signatures of resistance overcoming. Overall, even for such drastic demographic events, the resulting genetic signatures remain transients, and thus can only be captured by using time sample data in the appropriate (short) time window.

4.4 Temporal changes of demogenetic signatures

Among indices, we observe different temporal dynamics of genetic signatures through resistance overcoming. The fastest changes are observed for the genetic differentiation between S and R and the mean inbreeding coefficient, during the 50 first generations after the resistance overcoming. This is consist-

ent with the resistance overcoming of resistance RMlp7 in poplars by the poplar rust pathogen that led to drastic genetic changes in less than five years (Persoons et al., 2017). Changes in linkage disequilibrium and mean number of alleles are slower and detected within the 100 first generations. Finally, the slowest changes (and less sharp temporal signatures) are observed for the expected heterozygosity and the temporal genetic differentiation, with most of the detectable signal occurring between generations 50 and 200. Overall no peak of genetic signature occurs after generation 200, and only residual signal remains on population genetic indices.

For indices reflecting disequilibrium induced by the founder effect ($F_{ST}\,R-S$, $\bar{r}_D$, $Mean\,F_{IS}$), genetic changes stabilise rapidly, especially when gene flow is enhanced by a life cycle with host alternation. For other indices (temporal F_{ST}, H_E, $Mean\,LA$) and when it exists, the difference with the null model (*i.e.* without selection) persists for a much longer period (for at least 300 generations for $Mean\,LA$ and until the end of the simulated period for temporal F_{ST} and H_E). Following rapid adaption, we thus observe a return to genetic equilibrium in two steps. The first step involves the homogenisation of allele frequencies both within and between populations thanks to the convergence to Hardy-Weinberg and migration-drift equilibria. This occurs very fast because the modelled system is regularly subjected to the stabilising effect of sexual reproduction (Rouger et al., 2016). We hypothesise that this return would be slower for systems that deviate from this mode of reproduction, either because of an increased rate of clonality (Reichel et al., 2016) or because of selfing (Jullien et al., 2019). The second step is the return to allele numbers and frequencies expected under the migration-mutation-drift equilibrium. This evolution is slower because it relies on the progressive change in allelic states and potential recovery of the lost genetic diversity that happened during the upheaval caused by resistance overcoming.

Therefore, observing the first return to equilibrium requires the sampling of the population close to the time of resistance overcoming. Conversely, observing the second return to equilibrium requires a broader temporal sampling including the characterisation of the initial state. In the absence of this initial state, it is not possible to highlight these changes in allele frequencies at the virulence locus and neutral markers (Persoons et al., 2017).

The timing of the modelled events may appear rapid but is consistent with empirical observations. For example, in the poplar rust pathogen (*Melampsora larici-populina*) which alternates on larch every year, the overcoming of resistance RMlp7 in poplars led in 1994 to a strong genetic disequilibrium. This was followed by a quick return to Hardy-Weinberg equilibrium the following year and drastic changes in the population genetic structure occurring in less than four years (Persoons et al., 2017; Louet et al., 2021).

Overall, studies employing time-series remain rare compared to the amount of work focusing on the genetic analysis of one contemporary pathogen population (Buffalo and Coop, 2019, 2020; Pavinato et al., 2021). Analyses of time-series are mostly used to analyse the speed and timing of selection for life history traits (Le Rouzic et al., 2011, 2015), loci under positive selection (Bergland et al., 2014; Foll et al., 2015), or coevolution between host and parasite (Decaestecker et al., 2007; Gandon et al., 2008). We note that studies including neutral markers mainly use them to draw a statistical inference of loci under selection rather than to document the epidemic. In the rare cases where time-series are used

to study specifically neutral genetic evolution, the data typically exhibit a limited temporal range (*i.e.* two time points (Pavinato et al., 2021). However, as temporal data allow to trace the changes in allele frequency through time, the analysis of neutral markers can improve our inference and understanding of evolutionary processes (Dehasque et al., 2020; Feder et al., 2021; Saubin et al., 2022). In cases of rapid adaptation especially, we show here that the resulting genetic signatures may be very brief and require time samples around the selection event. Studies focusing on one or two time points may allow documenting only the second return to equilibrium (from the influence of evolutionary forces) and will fail to highlight transient dynamics, which provide the most relevant information regarding the demogenetic interplay. Conversely, epidemiological considerations lead to focus on the time when the invasion is detected, therefore when the genetic signature is the strongest. These considerations tend to neglect the initial state, as well as the return to a new equilibrium, which may nevertheless occur in a relatively short time scale. A key result of our study is to demonstrate that neutral markers can be used to uncover demogenetic processes due to selection events.

4.5 Concluding remarks

Our demogenetic analysis and new simulation framework enable us to study in depth the entire range of possible scenarios of resistance overcoming and to identify and discriminate between several eco-evolutionary dynamics using neutral genetic markers. To deal with the high amount of information from stochastic dynamics we performed clustering based on the time-series changes of population genetics indices. Using this method, we highlight three main types of genetic signatures of increasing magnitude that reflect different demographic processes through resistance overcoming: 1) small variations in population sizes, 2) a strong founder event, and 3) a strong founder event associated with a population bottleneck. We speculate that this simulation framework and our results pave the way for inferring demogenetic dynamics underpinning rapid adaptation using empirical genetic data.

Acknowledgements

We thank Alexis Sarda for insightful answers on the DTWCLUST package. This work was supported by grants from the French National Research Agency (ANR-18-CE32-0001, CLONIX2D project; ANR-11-LABX-0002-01, Cluster of Excellence ARBRE). Méline Saubin was supported by a PhD fellowship from INRAE and the French National Research Agency (ANR-18-CE32-0001, CLONIX2D project).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agapow, P. M. and Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, 1(1-2):101–102.

Agrios, G. N. (2005). *Plant pathology*. Elsevier edition.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., and Daszak, P. (2004). Emerging infectious diseases of plants : pathogen pollution , climate change and agrotechnology drivers. *Trends in Ecology and Evolution*, 19(10):536–544.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., and Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1):243–256.

Archie, E. A., Luikart, G., and Ezenwa, V. O. (2009). Infecting epidemiology with genetics: A new frontier in disease ecology. *Trends in Ecology and Evolution*, 24(1):21–30.

Arnaud-Haond, S., Duarte, C. M., Alberto, F., and Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, 16:5115–5139.

Barrett, R. D. H. and Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, 23(1):38–44.

Bell, G. (2017). Evolutionary rescue. *Annual Review of Ecology, Evolution, and Systematics*, 48:605–627.

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., and Petrov, D. A. (2014). Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLoS Genetics*, 10(11).

Bolker, B. M., Nanda, A., and Shah, D. (2010). Transient virulence of emerging pathogens. *Journal of the Royal Society Interface*, 7.

Bonneaud, C. and Longdon, B. (2020). Using evolutionary theory to understand the fate of novel infectious pathogens. *Science Society*, 21(e51374).

Brown, J. H. and Kodric-Brown, A. (1977). Turnover rates in insular biogeography: Effect of immigration on extinction. *Ecology*, 58:445–449.

Brown, J. K. M. and Tellier, A. (2011). Plant-parasite coevolution: Bridging the gap between genetics and ecology. *Annual Review of Phytopathology*, 49(1):345–367.

Buffalo, V. and Coop, G. (2019). *The linked selection signature of rapid adaptation in temporal genomic data*, volume 213.

Buffalo, V. and Coop, G. (2020). Estimating the genome-wide contribution of selection to temporal allele frequency change. *Proceedings of the National Academy of Sciences of the United States of America*, 117(34):20672–20680.

Burdon, J. J., Zhan, J., Barrett, L. G., Papaix, J., and Thrall, P. H. (2016). Addressing the challenges of pathogen evolution on the world's arable crops. *Phytopathology*, 106(10):1117–1127.

Carlson, S. M., Cunningham, C. J., and Westley, P. A. H. (2014). Evolutionary rescue in a changing world. *Trends in Ecology and Evolution*, 29(9):521–530.

Cornuet, J. M. and Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, 144(4):2001–2014.

Day, T. and Gandon, S. (2007). Applying population-genetic models in theoretical evolutionary epidemiology. *Ecology Letters*, 10:876–888.

Day, T. and Proulx, S. R. (2004). A general theory for the evolutionary dynamics of virulence. *The American naturalist*, 163(4).

Decaestecker, E., Gaba, S., Raeymaekers, J. A. M., Stoks, R., Van Kerckhoven, L., Ebert, D., and De Meester, L. (2007). Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature*, 450(7171):870–873.

Dehasque, M., Ávila-Arcos, M. C., Díez-del Molino, D., Fumagalli, M., Guschanski, K., Lorenzen, E. D., Malaspinas, A. S., Marques-Bonet, T., Martin, M. D., Murray, G. G. R., Papadopulos, A. S. T., Therkildsen, N. O., Wegmann, D., Dalén, L., and Foote, A. D. (2020). Inference of natural selection from ancient DNA. *Evolution Letters*, 4(2):94–108.

Dorken, M. E. and Eckert, C. G. (2001). Severely reduced sexual reproduction in northern populations of a clonal plant, *Decodon verticillatus* (Lythraceae). *Journal of Ecology*, 89(3):339–350.

Feder, A. F., Pennings, P. S., and Petrov, D. A. (2021). The clarifying role of time series data in the population genetics of HIV. *PLoS Genetics*, 17(1):1–10.

Fischhoff, I. R., Huang, T., Hamilton, S. K., Han, B. A., LaDeau, S. L., Ostfeld, R. S., Rosi, E. J., and Solomon, C. T. (2020). Parasite and pathogen effects on ecosystem processes : A quantitative review. *Ecosphere*, 11(5).

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., and Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, 484(7393):186–194.

Foll, M., Shim, H., and Jensen, J. D. (2015). WFABC: A Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources*, 15(1):87–98.

Gandon, S., Buckling, A., Decaestecker, E., and Day, T. (2008). Host-parasite coevolution and patterns of adaptation across time and space. *Journal of Evolutionary Biology*, 21(6):1861–1866.

Gladieux, P., Byrnes, E. J., Aguileta, G., Fisher, M. C., Heitman, J., and Giraud, T. (2011). *Epidemiology and evolution of fungal pathogens in plants and animals*.

Gokhale, C. S., Papkou, A., Traulsen, A., and Schulenburg, H. (2013). Lotka – Volterra dynamics kills the Red Queen : population size fluctuations and associated stochasticity dramatically change host-parasite coevolution. *BMC Evolutionary Biology*, 13(254):3–10.

Gonzalez, A., Ronce, O., Ferriere, R., and Hochberg, M. E. (2013). Evolutionary rescue: An emerging focus at the intersection between ecology and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1610):1–8.

Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., and Duplessis, S. (2011). The poplar-poplar rust interaction: Insights from genomics and transcriptomics. *Journal of Pathogens*, 2011:1–11.

Halkett, F., Simon, J. C., and Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology and Evolution*, 20.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Johnson, R. (1984). A critical analysis of durable resistance. *Annual Review of Phytopathology*, 22.

Jullien, M., Navascués, M., Ronfort, J., Loridon, K., and Gay, L. (2019). Structure of multilocus genetic diversity in predominantly selfing populations. *Heredity*, 123(2):176–191.

Laine, A. L. and Barrès, B. (2013). Epidemiological and evolutionary consequences of life-history trade-offs in pathogens. *Plant Pathology*, 62:96–105.

Le Rouzic, A., Hansen, T. F., Gosden, T. P., and Svensson, E. I. (2015). Evolutionary time-series analysis reveals the signature of frequency-dependent selection on a female mating polymorphism. *American Naturalist*, 185(6):E182–E196.

Le Rouzic, A., Houle, D., and Hansen, T. F. (2011). A modelling framework for the analysis of artificial-selection time series. *Genetics Research*, 93(2):155–173.

Leroy, T., Lemaire, C., Dunemann, F., and Le Cam, B. (2013). The genetic structure of a *Venturia inaequalis* population in a heterogeneous host population composed of different *Malus* species. *BMC Evolutionary Biology*, 13(1).

Louet, C., Saubin, M., Andrieux, A., Persoons, A., Gorse, M., Pétrowski, J., Fabre, B., De Mita, S., Duplessis, S., Frey, P., and Halkett, F. (2021). A point mutation and large deletion at the candidate avirulence locus *AvrMlp7* in the poplar rust fungus correlate with poplar RMlp7 resistance breakdown. *Molecular Ecology*, (February):1–12.

Luchman, J. (2022). *domir: Tools to support relative importance analysis*.

McCauley, D. E. (1991). Genetic consequences of local population extinction and recolonization. *Trends in Ecology and Evolution*, 6(1):5–8.

McDonald, B. A. (2004). Population genetics of plant pathogens. *The Plant Health Instructor*.

McDonald, B. A. and Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, 40:349–379.

Milgroom, M. G. and Peever, T. L. (2003). Population biology of plant pathogens: The synthesis of plant disease epidemiology and population genetics. *Plant Disease*, 87(6):608–617.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3):583–590.

Nei, M. and Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics*, 98(3):625–640.

Nielsen, R. and Signorovitch, J. (2003). Correcting for ascertainment biases when analyzing SNP data: Applications to the estimation of linkage disequilibrium. *Theoretical Population Biology*, 63(3):245–255.

Otto, S. P. and Whitlock, M. C. (1997). The probability of fixation in populations of changing size. *Genetics*, 146(2):723–733.

Parsons, T. L., Lambert, A., Day, T., and Gandon, S. (2018). Pathogen evolution in finite populations : slow and steady spreads the best. *Journal of royal society*, 15.

Pavinato, V. A. C., De Mita, S., Marin, J. M., and de Navascués, M. (2021). Joint inference of adaptive and demographic history from temporal population genomic data. *bioRxiv*, pages 1–18.

Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., and Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, 26(7):1902–1918.

Pink, D. and Puddephat, I. (1999). Deployment of disease resistance genes by plant transformation - a 'mix and match' approach. *Trends in Plant Science*, 4(2):71–75.

Pollak, E. (1983). A new method for estimating the effective population size from allele frequency changes. *Genetics*, 104(3):531–548.

R (2018). R Core Team, R: A language and environment for statistical computing.

Reichel, K., Masson, J. P., Malrieu, F., Arnaud-Haond, S., and Stoeckel, S. (2016). Rare sex or out of reach equilibrium? The dynamics of FIS in partially clonal organisms. *BMC Genetics*.

Rouger, R., Reichel, K., Malrieu, F., Masson, J. P., and Stoeckel, S. (2016). Effects of complex life cycles on genetic diversity: Cyclical parthenogenesis. *Heredity*, 117(5):336–347.

Sardá-Espinosa, A. (2019). Time-series clustering in R using the dtwclust package. *R Journal*, 11(1):1–22.

Saubin, M., De Mita, S., Zhu, X., Sudret, B., and Halkett, F. (2021). Impact of ploidy and pathogen life cycle on resistance durability. *Peer Community Journal*, 1(e8):1–31.

Saubin, M., Louet, C., Bousset, L., Fabre, F., Frey, P., Fudal, I., Grognard, F., Hamelin, F., Mailleret, L., Stoeckel, S., Touzeau, S., Petre, B., and Halkett, F. (2022). Improving sustainable crop protection using population genetics concepts. *Molecular Ecology*, 00:1–11.

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., Mcroberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature ecology evolution*, 3(3):430–439.

Stoeckel, S. and Masson, J. P. (2014). The exact distributions of FIS under partial asexuality in small finite populations with mutation. *PLoS ONE*, 9(1).

Stukenbrock, E. H. and McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, 46:75–100.

Stukenbrock, E. H. and McDonald, B. A. (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions*, 22(4):371–380.

Thrall, P. H. (1998). Metapopulation collapse: The consequences of limited gene flow in spatially structured populations. In *Modeling Spatiotemporal Dynamics in Ecology,*, pages 83–104. Springer–Verlag.

Thrall, P. H., Barrett, L. G., Dodds, P. N., and Burdon, J. J. (2016). Epidemiological and evolutionary outcomes in gene-for-gene and matching allele models. *Frontiers in Plant Science*, 6(1084).

Tobin, P. C. (2015). Ecological consequences of pathogen and insect invasions. *Forest pathology and entomology*, 1:25–32.

van Rossum, G. (1995). Python tutorial, Technical Report CS-R9526. *CWI*.

Wahlund, V. S. (1928). Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet. *Hereditas*, 11:65–106.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16.

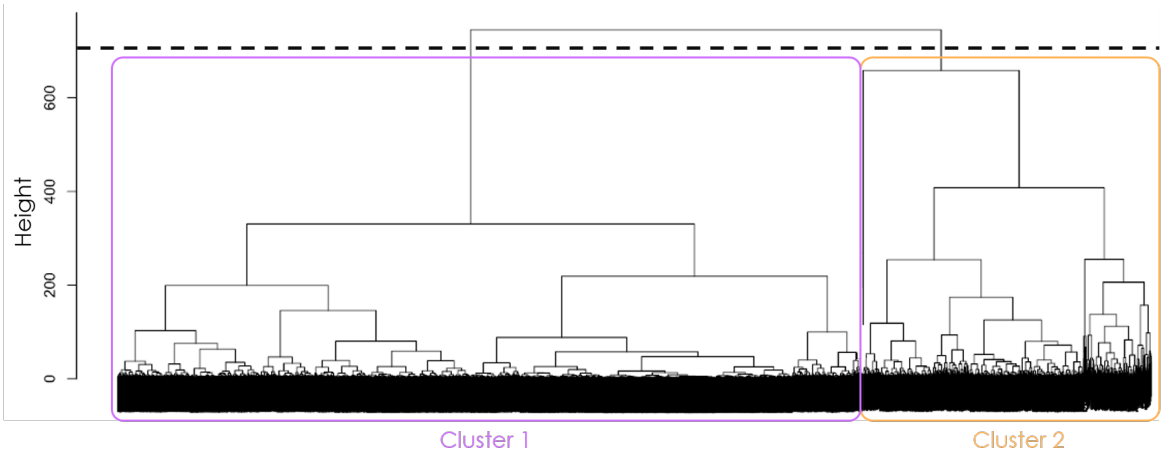
Wright, S. (1949). *Adaptation and selection*, chapter Adaptation, pages 365–389. Princeton Univ. Press, Princeton, NJ.

Wright, S. (1978). *Evolution and the genetics of populations - Variability within and among natural populations*. The University of Chicago Press.

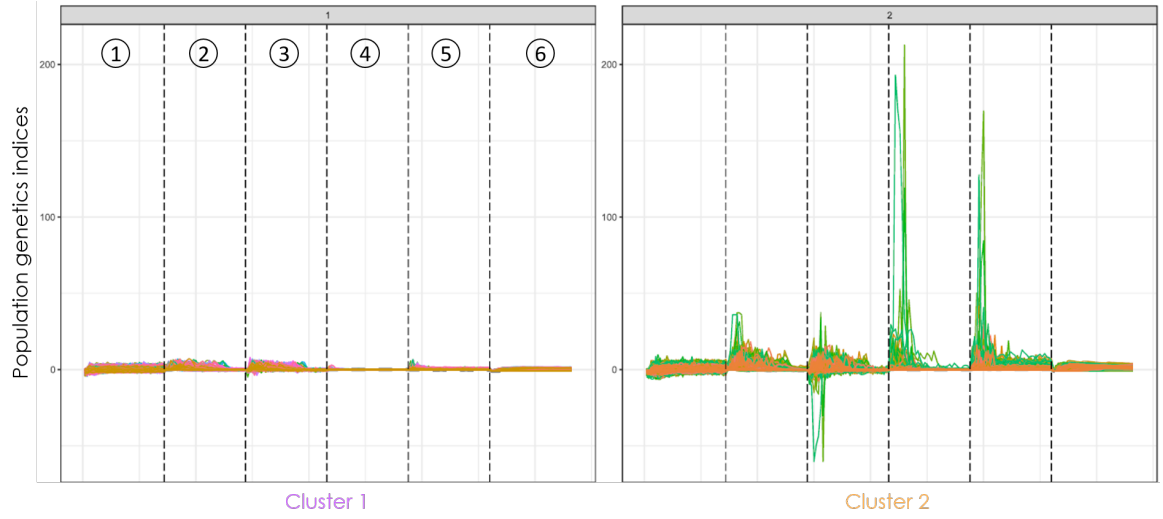
Zhan, J., Thrall, P. H., Papaïx, J., Xie, L., and Burdon, J. J. (2015). Playing on a pathogen's weakness: Using evolution to guide sustainable plant disease control strategies. *Annual Review of Phytopathology*, 53(1):19–43.

Živković, D., John, S., Verin, M., Stephan, W., and Tellier, A. (2019). Neutral genomic signatures of host-parasite coevolution. *BMC Evolutionary Biology*, 19(1):1–11.

Appendix

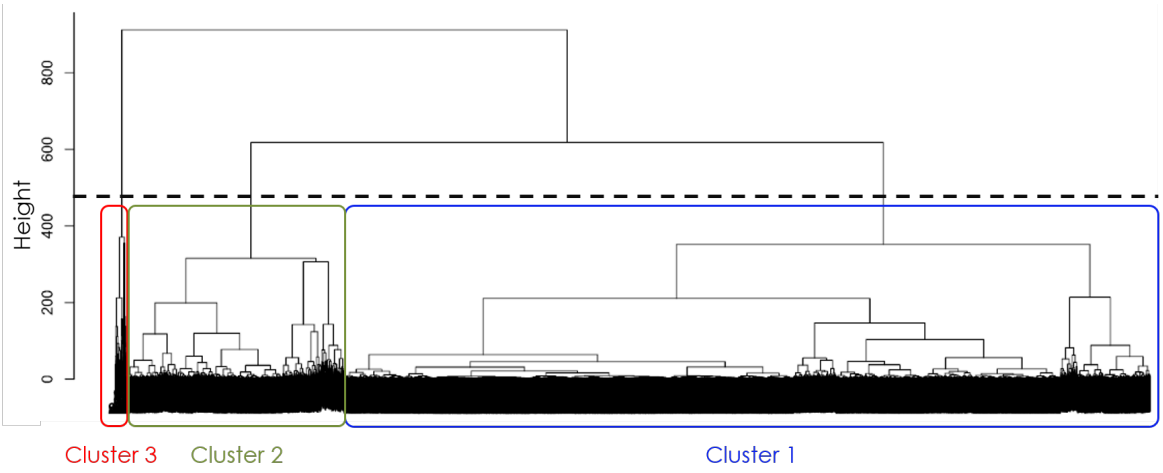


(a) Clustering of simulations

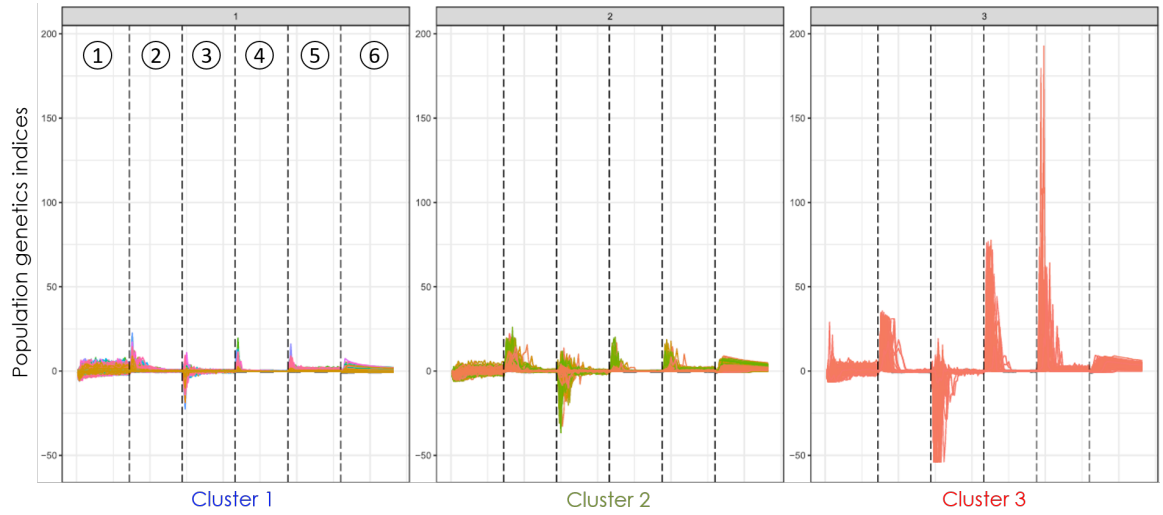


(b) Temporal evolution of population genetics indices

Figure S1: Clustering of simulations without host alternation. The clustering is based on the temporal evolution of six population genetics indices. (a) represents the dendrogram obtained from the clustering analysis. Distances between simulations were calculated based on the dynamic time warping distance, with the norm 'Eucliean distance' and the step pattern 'symmetric 2'. (b) represents for each cluster the concatenation of the temporal evolution of six normalised populations genetics indices, in the following order: (1) $\beta\, of Pareto$, (2) $\bar{r}_D$, (3) $Mean\ F_{IS}$, (4) $Variance\ F_{IS}$, (5) F_k, (6) $Temporal\ F_{ST}$.



(a) Clustering of simulations



(b) Temporal evolution of population genetics indices

Figure S2: Clustering of simulations with host alternation. The clustering is based on the temporal evolution of six population genetics indices. (a) represents the dendrogram obtained from the clustering analysis. Distances between simulations were calculated based on the dynamic time warping distance, with the norm 'Eucliean distance' and the step pattern 'symmetric 2'. (b) represents for each cluster the concatenation of the temporal evolution of six normalised populations genetics indices, in the following order: (1) $\beta\, of Pareto$, (2) $\bar{r}_D$, (3) $Mean\ F_{IS}$, (4) $Variance\ F_{IS}$, (5) F_k, (6) $Temporal\ F_{ST}$.

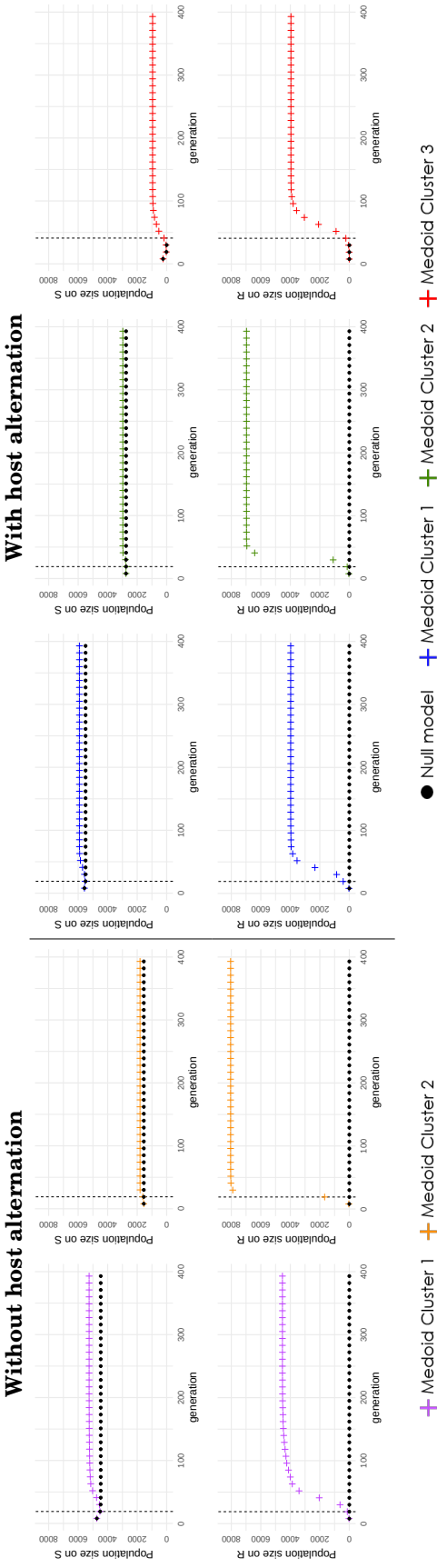
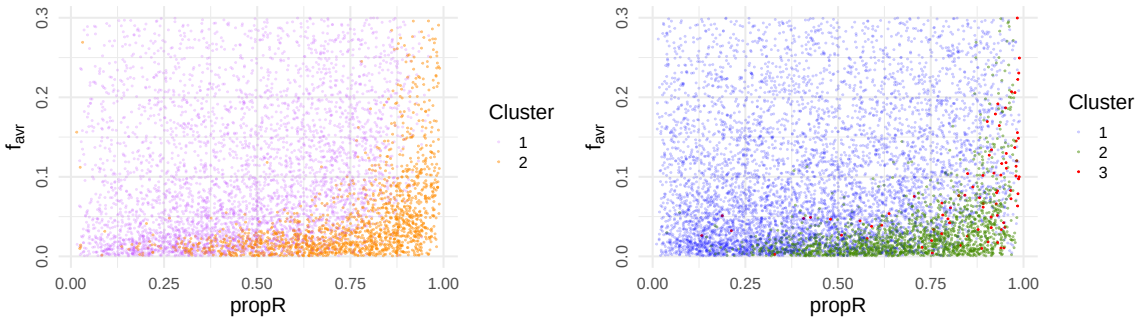


Figure S3: Comparison of the temporal evolution of population sizes, between medoid simulations and null models. The plotted results correspond to the medoid dynamic of each cluster, with (‘Medoid’, cross shapes) or without (‘Null model’, round shapes) virulent allele, that is, with or without resistance overcoming.



(a) Without host alternation

(b) With host alternation

Figure S4: Distribution of $propR$ and f_{avr} input parameters for each simulations of both life cycles. Each point represents a simulation of the random simulation design that led to resistance overcoming, coloured by its cluster assignment.

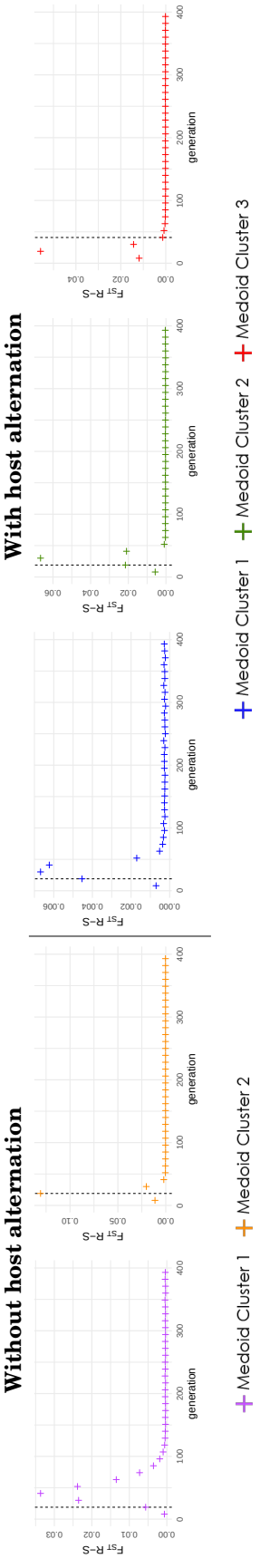


Figure S5: Temporal evolution of F_{ST} $R-S$. The plotted results correspond to the medoid dynamics for each cluster and each life cycle. For each of the plotted medoid, the vertical dashed line corresponds to the generation of resistance overcoming.

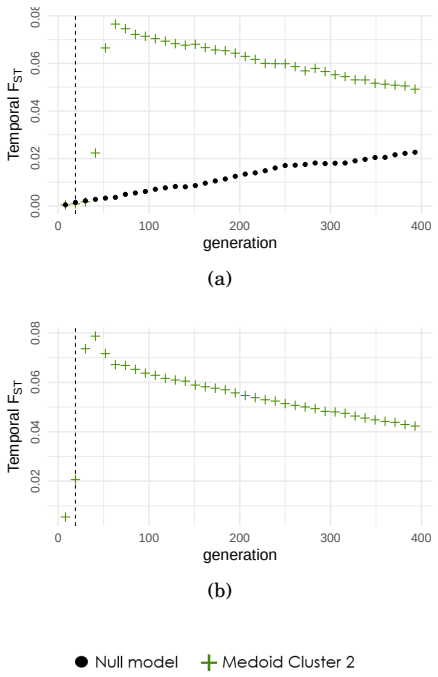


Figure S6: Comparison of the medoid dynamics of $Temporal\ F_{ST}$ on (a) S and (b) R for Cluster 2 with host alternation. Subfigure (a) is already presented in the main document (Figure 6, with host alternation, Cluster 2).

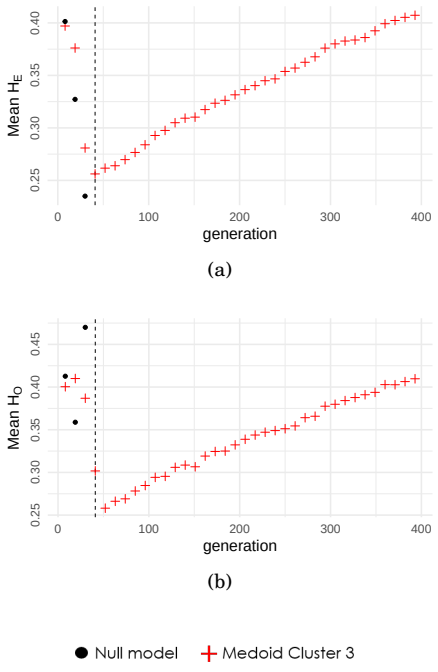


Figure S7: Comparison of the medoid dynamics of (a) $Mean\ H_E$ and (b) $Mean\ H_O$ on S for Cluster 3 with host alternation. Subfigure (a) is already presented in the main document (Figure 6, with host alternation, Cluster 3).

3.2 Conclusion en regard du contournement de la résistance RMlp7 par *Melampsora larici-populina*

À travers cette analyse, nous avons construit une typologie détaillée des scénarios évolutifs et démographiques des agents pathogènes lors d'un contournement de résistance. Les signatures génétiques mises en évidence permettent notamment de distinguer trois grands types de processus démographiques associés à un contournement de résistance : 1) de très faibles variations de tailles de populations, 2) un événement fondateur fort sur l'hôte résistant et 3) un événement fondateur fort sur l'hôte résistant associé à un goulot d'étranglement de la population sur l'hôte sensible.

Nous avons mis en évidence l'impact de la proportion d'hôtes résistants sur les signatures génétiques résultant d'un contournement. Nous observons de très fortes variations génétiques lorsque cette proportion est élevée. Cette forte proportion d'hôtes résistants peut ainsi expliquer le changement rapide de la structure génétique des populations de Mlp lors du contournement de la résistance RMlp7 et notamment le remplacement du groupe génétique « Fossile » par un nouveau groupe génétique « Cultivé » (FIGURE 1.5).

Pour confronter les attendus théoriques aux données empiriques, nous avons calculé les mêmes indices de génétique des populations sur 22 échantillonnages annuels de Mlp prélevés dans un même lieu (Amance, France) entre 1989 et 2021 (FIGURE 3.1). Cette chronoséquence représente 504 individus échantillonnés sur des peupliers sensibles. La taille des populations échantillonnées varie de 5 à 58 individus, les populations de moins de 5 individus ayant été retirées de l'analyse.

L'évolution génétique neutre chez Mlp montre une augmentation du F_{ST} temporel à partir de 1994, l'année du contournement de la résistance RMlp7. Sur cette chronoséquence de 30 ans, la valeur de maximale du F_{ST} temporel est de 0.07. Cet ordre de grandeur correspond à une signature génétique caractéristique d'un fort effet de fondation sur le compartiment résistant lors du contournement. Ce résultat est cohérent avec le remplacement de population rapide observé au cours de ce contournement (FIGURE 1.5). Nous observons également dans les données empiriques une diminution puis une remontée rapide de l'hétérozygotie théorique (Moyenne H_E), précédée d'une légère diminution du nombre moyen d'allèles dans la population. Ces variations indiquent un léger goulot d'étranglement de la population de Mlp précédant le contournement de résistance. Ces signatures ne sont cependant pas aussi importantes que les cas extrêmes observés dans les simulations et causées par des évènements de sauvetage évolutif.

Enfin, l'augmentation particulièrement importante de l'indice de fixation F_{IS} au moment du contournement de la résistance RMlp7 ($F_{IS} > 0.15$ en 1994, FIGURE 3.1) indique une juxtaposition d'individus fortement différenciés sur les peupliers sensibles. Ce signal ne correspond pas aux variations de F_{IS} retrouvées dans notre modèle pour un agent pa-

thogène avec alternance d'hôte. En effet, d'après les résultats du modèle, nous nous attendons à ce que le signal lié au mélange des génotypes soit très rapidement atténué par la reproduction sexuée sur un hôte commun qui ferait disparaître ces différenciations. Nous touchons ici à une limite du modèle : nous modélisons trois compartiments d'hôtes uniquement (peupliers sensible, peupliers résistants et mélèzes) en négligeant les structures spatiales plus complexes que l'on peut rencontrer dans les paysages agricoles. Les variations de F_{IS} observées dans les données empiriques sont cohérentes en regard de la spatialisation des plantations de peupliers et de mélèzes. Cela reflète l'existence de plusieurs populations différenciées sur le territoire et une invasion des peupliers résistants qui ne s'est pas nécessairement faite à la même vitesse sur l'ensemble des peupleraies.

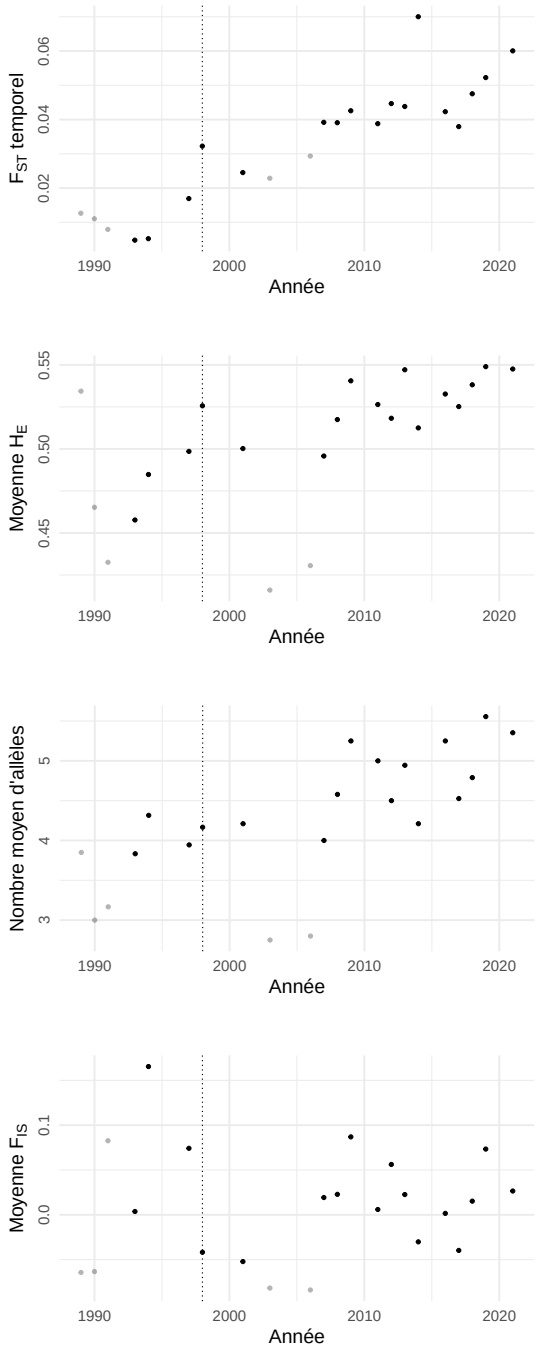


FIGURE 3.1 – Évolution de quatre indices de génétique des populations calculés à partir d'isolats de Mlp échantillonnés dans un même lieu (Amance, France) entre 1989 et 2021 sur des peupliers sensibles. La ligne verticale correspond à l'année de détection des premiers isolats virulent 7 dans ce lieu ; Les points noirs correspondent aux populations composées d'au moins 10 individus ; Les points gris correspondent aux populations composées de moins de 10 individus et dont les résultats sont à prendre avec précaution. Les populations de moins de cinq individus ont été retirées de l'analyse.

Chapitre 4

Inférence d'un scénario de contournement

Le chapitre précédent nous a permis d'expliquer l'évolution des structures génétiques observées chez Mlp lors du contournement de la résistance RMlp7. Je souhaite maintenant déterminer si le modèle développé peut nous permettre de dépasser le cadre de la compréhension théorique du système, pour aller jusqu'à une inférence des paramètres écologiques et démographiques ayant régi le contournement de la résistance RMlp7.

L'objectif de ce chapitre est de développer une méthode d'inférence par analyse ABC à partir de données temporelles. Nous savons d'après les chapitres précédents que les signatures génétiques d'un contournement de résistance peuvent être très brèves, avec un retour rapide à l'équilibre. Je souhaite donc en particulier déterminer à quel point la fréquence et la régularité des échantillonnages dans le temps impactent la qualité de l'inférence.

4.1 Article 4. Approximate Bayesian Computation applied to time series of population genetic data disentangles rapid genetic changes and demographic variations in a pathogen population

À partir des signatures génétiques du contournement, je souhaite inférer dans un premier temps le cycle de vie de l'agent pathogène : avec ou sans alternance d'hôte. Je souhaite ensuite inférer six paramètres écologiques et démographiques régissant un contournement de résistance : 1) La fréquence initiale d'allèles virulents dans la population, 2) Le taux de migration de l'agent pathogène entre hôtes sensibles et résistants, 3) Le taux de croissance de la population d'agents pathogènes, 4) Le taux de mortalité lors de la migration annuelle sur l'hôte alternant, 5) La capacité de charge cumulée des hôtes sensibles et résistants, 6) La proportion d'hôtes résistants dans le paysage.

J'adapte le modèle développé précédemment aux données empiriques dont je dispose : je considère cette fois l'évolution de 23 marqueurs neutres caractérisés par 20 états alléliques possibles et au taux de mutation de 10^{-3}, à la manière de marqueurs microsatellites. Les populations échantillonnées dans les simulations sont maintenant composées de 30 individus tirés au hasard sur chaque compartiment. Je développe un cadre d'inférence par la méthode ABC à partir de ce modèle. Je détermine dans un premier temps l'identifiabilité théorique de chaque paramètre en étudiant notamment l'influence de la structure temporelle de l'échantillonage sur la puissance de l'analyse. J'applique ensuite cette méthode aux données empiriques dont je dispose sur le contournement de la résistance RMlp7, en adaptant la structure des échantillons simulés à la structure des données empiriques.

Ces résultats sont ici présentés sous la forme d'un article de recherche en préparation intitulé « Approximate Bayesian Computation applied to time series of population genetic data disentangles rapid genetic changes and demographic variations in a pathogen popu-

lation » et qui sera soumis à la revue MOLECULAR ECOLOGY en octobre 2022.

Approximate Bayesian Computation applied to time series of population genetic data disentangles rapid genetic changes and demographic variations in a pathogen population

Méline Saubin[1], Aurélien Tellier[2], Solenn Stoeckel[3], Axelle Andrieux[1] and Fabien Halkett[1]

[1] Université de Lorraine, INRAE, IAM, F-54000 Nancy, France

[2] Professorship for Population Genetics, Technical University of Munich, Freising, Germany

[3] INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, F-35653 Le Rheu, France

1 Introduction

Adaptation can occur at remarkably short timescales in natural populations, leading to drastic changes in genotype frequencies and phenotypes over a few generations only (Buffalo and Coop, 2019). This rapid pace of adaptation has motivated the use of temporal data to understand neutral (Prout, 1954; Wallace, 1956; Nei and Tajima, 1981; Pollak, 1983; Mueller et al., 1985b; Waples, 1989; Wang and Whitlock, 2003) and under-selection genetic evolution over time (Dobzhansky, 1943; Fisher and Ford, 1947; Kettlewell, 1958, 1961; Mueller et al., 1985a). However, such time-series studies remain rare compared to the amount of work focusing on one contemporary sample to trace back its genetic history (Buffalo and Coop, 2019, 2020; Pavinato et al., 2021). Temporal data allow to track the changes in allele frequency through time, and therefore lead to a better understanding of evolutionary processes (Dehasque et al., 2020; Feder et al., 2021; Saubin et al., 2022). In cases of rapid adaptation especially, the resulting genetic changes may be transient and require time sampling around the selection event to be highlighted (Article 3).

The inference of demographic parameters can allow understanding how evolutionary forces (especially genetic drift and selection) interact and shape the genetic trajectories of populations during rapid adaptation (Bergland et al., 2014; Živković et al., 2019). However, the

difficulties in obtaining the likelihood of models including both demography, selection, and genetic drift (Pavinato et al., 2021) lead to consider alternative approaches relying on simulations (Bazin et al., 2010; Laval et al., 2019). With the advent of computational approaches, Approximate Bayesian Computation (ABC) has become a standard approach for genetic analyses aimed at tracing back the evolutionary history of populations (Rosenberg and Nordborg, 2002; Bazin et al., 2010; Cornuet et al., 2010; Estoup and Guillemaud, 2010). These methods, coupled with a coalescent simulator, allow inferring evolutionary scenarios and demographic parameters, by modelling the genealogy of the samples (Whitlock and Barton, 1997). The coalescent is computationally very efficient but relies on strong modelling assumptions (Rosenberg and Nordborg, 2002). As such, coalescent-based ABC approaches suffer from two limitations. First, there is a gap between the models developed so far and the actual complexity of biological scenarios. Different coalescent models have been developed to account for fluctuating population sizes (Sjödin et al., 2005) or partially clonal species (Orive, 1993; Hartfield, 2021). Yet these specificities has been considered in isolation. To date, few models integrate both rapid demographic fluctuations and complex life cycles (Tellier and Lemaire, 2014). Second, coalescent-based inference methods are powerful only when coalescent events occur at the same time scale as the process being considered, such as past demographic events. This means that demographic events happening too fast compared to the coalescent time scale (in units of the effective population size N_e) cannot be inferred. For example, under antagonistic coevolution, changes in the demography of host and pathogen are too fast for a coalescent analysis, but tracking allele frequencies and nucleotide diversity forward-in-time is informative about this process (Živković et al., 2019). With a few exceptions (*e.g.* De Mita and Siol, 2012; Foll et al., 2015), time samples are not considered in such frameworks. Rapid adaptation processes are associated with transient genetic changes that can be well illuminated by considering step-by-step modelling algorithms forward-in-time. The study of the rapid adaptation of pathogens would therefore benefit from being considered with forward models (Foll et al., 2015). Yet they consider so far relatively simple demography. In the case of more complex scenarios, the first step is to decipher how a selective event shape population structure at neutral loci (Hoban et al., 2016). Here we propose a new ABC framework that couples a forward and individual-based model with temporal genetic data disentangle genetic changes and demographic variations in a case of rapid adaptation.

Understanding and inferring the evolutionary trajectories of populations is of major interest to evolutionary biologists but it can also have practical applications for population management. This is especially the case in agriculture where the need to control pathogen populations is paramount (Parat et al., 2016). Pathogens can induce disease outbreaks dev-

astating human-managed ecosystems (Anderson et al., 2004; Tobin, 2015; Savary et al., 2019). Understanding the evolution of pathogens is therefore crucial for developing effective disease management strategies (Bonneaud and Longdon, 2020). The rapid adaptation of pathogen populations (McDonald and Linde, 2002; Saubin et al., 2021) and the high stochasticity in pathogen evolutionary trajectories (Parsons et al., 2018) make this endeavour extremely difficult.

The rapid adaptation of pathogen populations in agrosystems results from the tremendous selection pressures exerted by modern agricultural practices (Zhan et al., 2015; Stukenbrock and McDonald, 2008). To counter plant disease outbreaks, breeders develop resistant plant genotypes. These genetic resistances are most often deployed across large spatial scales (Zhan et al., 2015; Rimbaud et al., 2021). These agricultural practices break eco-evolutionary feedbacks that maintain the polymorphism observed in natural host-parasite systems (Stukenbrock and McDonald, 2008; Brown and Tellier, 2011). As such, these practices weaken the sustainability of plant genetic resistances in favouring the emergence and spread of virulent (*i.e.* resistance-adapted) pathogens (Rimbaud et al., 2021; Saubin et al., 2022). Therefore, the outcome of plant genetic resistance deployment is often a resistance overcoming event, *i.e.* the failure of the host plant to remain resistant to the pathogen. This results in the spread of virulent pathogens on resistant hosts (Johnson, 1984; Pink and Puddephat, 1999; Brown and Tellier, 2011; Burdon et al., 2016). On the pathogen side, an event of resistance overcoming can translate into a strong selective sweep with the intense and unidirectional selection causing drastic demographic changes for the pathogen population (Burdon et al., 2016; Persoons et al., 2017; Saubin et al., 2021). Such rapid adaptation can lead to specific temporal genetic signatures at neutral loci, depending on the evolutionary scenario ruling the change in population sizes (Article 3).

Here, we propose to use ABC based on temporal genetic evolution to unravel the evolutionary scenarios following rapid and contemporaneous adaptation. We apply our inferential framework to time series data and evaluate the added value of considering the full trajectory following forward simulations compared to few time samples. We test the accuracy of several temporal sampling designs in inferring model parameter values. Last, we apply our ABC inferential framework to empirical data describing the population genetics changes of the poplar rust pathogen following a major event of resistance overcoming (Persoons et al., 2017; Louet et al., 2021).

2 Materials and methods

2.1 Simulation model

The rapid adaptation we model is a resistance overcoming event that is monitored through time. We use an individual-based, forward-time and non-spatial demogenetic model, designed for diploid individuals (Saubin et al., 2021, Article 3). This model couples population dynamics and population genetics to follow the evolutionary trajectory of different genotypes at the selected locus and at neutral genetic markers. The model is implemented in Python (version 3.7, van Rossum, 1995) and Numpy (Harris et al., 2020). We consider life cycles commonly found in temperate pathogen species, with seasonal variation in reproductive mode. These pathogens switch from clonal reproduction during the epidemic phase to sexual reproduction occurring once a year, in winter (Agrios, 2005). This general life cycle is adjusted in two variations to take into account that the sexual reproduction takes place either on the same host plant as for the epidemic phase or on an alternate host (usually a different species). These life cycles are named with or without host alternation, respectively. Without host alternation, the model represents the evolution in time of a population of pathogens on two static host compartments: a resistant compartment (R) and a susceptible compartment (S). Each compartment has a fixed carrying capacity for the pathogen population, K_R and K_S, respectively. With host alternation, the alternate host compartment (A) is added, to account for the sexual reproduction occurring on another species. This static compartment is assumed to be larger than the two other compartments, with a fixed carrying capacity K_A. In the following, we refer to the three compartments as S (susceptible hosts), R (resistant hosts) and A (alternate hosts). The successive steps modelled through generations are presented Figure 1, and detailed in Saubin et al. (2021). We simulate the evolution at neutral loci and at a selected locus responsible for the virulence (qualitative trait) of pathogen individuals. Evolution at neutral loci is set to suit classical population genetic analyses based on microsatellite markers: 23 loci with a mutation rate of 10^{-3} and a maximum of 20 allelic states.

Each simulation starts with genotypes randomly drawn from the 20 possible alleles followed by a burn-in period with a population of constant size $N = K_S$. During the burn-in period, the pathogen population only evolves on S. To reach the genetic drift and mutation equilibrium from the initial state, the burn-in period is set to $2N$ generations. We build a random simulation design of 150,000 simulations, with the following parameter values drawn randomly from defined prior distributions (Table 1): 1) the proportion f_{avr} of virulent alleles introduced after the burn-in period in the pathogen population, reflecting initial standing genetic variation; 2) the life cycle of the pathogen ($Cycle$, with or without host alternation); 3) the

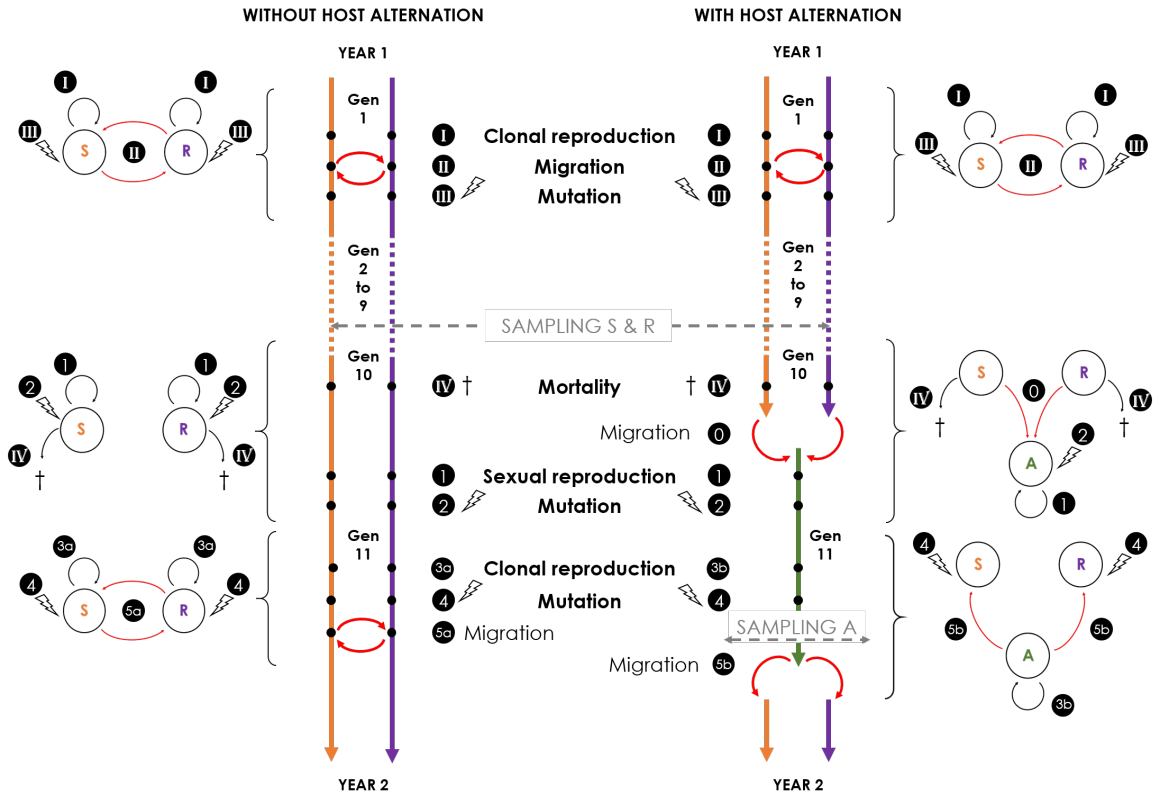


Figure 1: Modelling steps for each simulated year with the three S, R, and A host compartments. Each year is composed of 11 generations. During the clonal phase (generation 1 to 9), each generation is composed of three steps identical between both life cycles: (I) clonal reproduction; (II) migration of a proportion mig of each population between R and S; (III) mutation at all neutral markers with a mutation rate of 10^{-3}. At the end of the clonal phase, the pathogen overwinter as a dormant stage and is subjected to (IV) mortality of a proportion τ of each population. Then, the sexual phase (generations 10 and 11) differs depending on the life cycle: (0) represents the migration of all individuals from R and S towards A; (1) sexual reproduction; (2) and (4) mutation of all neutral markers with a mutation rate of 10^{-3}; (3a) and (3b) clonal reproduction; (5a) migration of a proportion mig of each population between R and S; (5b) migration of all individuals from A towards R and S. A sampling takes place every year at the end of generation 9 on S and R and at generation 11 before the migration event (5b) on A.

migration rate (mig) of the pathogen population; 4) the growth rate (r) of the pathogen population; 5) the annual mortality rate of the pathogen population before the sexual reproduction (τ); 6) the cumulative carrying capacity of S and R ($K = K_R + K_S$); and 7) the proportional size of R ($propR = \frac{K_R}{K}$).

Each simulation is run for 400 generations (with 11 generations per year). We retain only simulations for which the frequency of virulent allele in the population on S at the end of the simulation is above the threshold of 60%. This threshold is chosen to ensure that only simulations effectively leading to the resistant host invasion and resistance overcoming are kept. Of the 150,000 simulations produced from the random simulation design, 106,058 lead

Table 1: Input parameters and their range of variations for the random simulation design run for 400 generations.

Variable	Description	Distribution	Interval
f_{avr}	Initial frequency of the virulent allele in the pathogen population	Log-uniform	[log(0.0005); log(0.3)]
$Cycle$	Life cycle of the pathogen	Binomial	Without or with host alternation, probability 0.5
mig	Migration rate between R and S	Log-uniform	[log(0.001); log(0.2)]
r	Growth rate of the pathogen population	Uniform	[1.1; 2]
τ	Mortality rate during the annual migration	Uniform	[0.5; 1[
K	Cumulative carrying capacity of susceptible and resistant hosts	Uniform	[1,000; 20,000]
$propR$	Proportion of resistant hosts in the landscape	Uniform	]0.01; 0.99[

to resistance overcoming and are used for the following analyses (50,547 with host alternation, and 55,511 without host alternation).

2.2 Summary statistics

Sampling represents a random draw of n individuals from a host compartment, with the sample size $n = 30$. The sampling occurs each year at the end of generation 9 on S and R, and additionally - for the life cycle with host alternation - at generation 11 on A before the redistribution to S and R (Figure 1). Sampled individuals are not removed (sampling with replacement) so that their genotypes could contribute to subsequent generations.

To summarise the observed and simulated datasets, we compute the following statistics for each sample of individuals: (i) the proportion of virulent individuals (P_{Vir}); (ii) the genotypic diversity estimated by Pareto's β (β_P); (iii) the proportion of unique genotypes ($R = \frac{G-1}{n-1}$ with G the number of unique genotypes and n the number of sampled individuals); (iv) the genetic diversity estimated by Simpson's index (S); (v) the genetic diversity estimated by Shannon-Wiener's index (S); (vi) the mean expected heterozygosity over all loci (MH_E); (vii) the variance of the expected heterozygosity over all loci (VH_E); (viii) the mean number of alleles per locus (MLA); (ix) the multilocus index of linkage disequilibrium ($\bar{r}_D$); the population differentiation index between populations on R and S sampled at the same generation ($F_{ST}R - S$); (x) the population differentiation index between the initial population (on S after burn-in) and the sampled population (TF_{ST}) (Table 2). All statistics are implemented along with the model. Summary statistics are calculated for each population, independently of the sampled compart-

ment (S, R or A), except for P_{Vir} which is not recorded for populations sampled on R (because all individuals living on the resistant host are virulent), and for $F_{ST}R-S$ which is calculated from populations on S and R from the same generation. In the following analyses, summary statistics from different compartments are considered as distinct summary statistics. Because of the temporal sampling, each summary statistic is recorded for multiple generations, which constitutes a time series of summary statistics.

Two methods are used to perform the analyses: 1) Summary statistics from different generations are considered as distinct summary statistics and hence all generations along the time series are taken into account (hereafter referred to as 'Complete summary statistics'); 2) We extract the mean, variance, minimum and maximum values along each time series, and considered these four values as the distinct set of summary statistics for each population genetic index (hereafter referred to as 'Wrap-up summary statistics').

Table 2: Description of population genetics indices used as summary statistics in the ABC analyses.

Index	Description	Reference
P_{Vir}	Proportion of virulent individuals	
β_P	Genotypic diversity index	Arnaud-Haond et al., 2007
R	Proportion of unique genotypes	
S	Simpson index of genetic diversity	
SW	Shannon-Wiener index of genetic diversity	
MH_E	Mean expected heterozygosity overall loci	Nei, 1978
VH_E	Variance of the expected heterozygosity overall loci	
MLA	Mean number of alleles by locus	Nielsen and Signorovitch, 2003
$\bar{r}_D$	Linkage disequilibrium index	Agapow and Burt, 2001
$F_{ST}\ R-S$	Population differentiation between populations on R and S	Wright, 1949, 1978
TF_{ST}	Population differentiation between the initial population (after burn-in) and the sampled population	

2.3 Approximate Bayesian Computation

We use the R package ABC (Csilléry et al., 2012) to perform an ABC analysis from the simulated data. This analysis involves two steps: model choice and parameter estimation. The model choice aims to distinguish between the two simulated life cycles, with and without host alternation. The parameter estimation aims to estimate each of the six quantitative parameters of the chosen simulation model (Table 1).

The model choice procedure is based on a weighted multinomial logistic regression (Fagundes et al., 2007) computed on 1% of the simulated datasets for which ϵ, the Euclidean distance between the summary statistics of the observed dataset and the simulated datasets, is the smallest (Beaumont et al., 2002). Bayes factors are calculated as the ratio of the posterior probabilities for the tested models (Kass and Raftery, 1995).

For the parameter estimation procedure, we use the 50,547 simulations obtained under the

model with host alternation. We estimate posterior distributions (mode and 95% confidence intervals) of each parameter by applying the neural networks regression method (Blum and François, 2010) implemented in the R package ABC based on the 1% of simulations closest to the observed data.

Cross-validation on simulated data

To assess the validity of the method, cross-validations are performed at both steps of the analysis. A leave-one-out cross-validation for the model choice is performed on 500 randomly drawn simulations from each of the two models: with and without host alternation. From the Bayes factors obtained, the probabilities to re-estimate the true model are calculated for each life cycle (*i.e.* the model that is indeed used for the randomly drawn simulation). Additionally, we perform Principal Component Analyses (PCA) from the values of summary statistics obtained from the two life cycles. We display envelopes containing 95% of the simulations. A leave-one-out cross-validation for the parameter estimation is performed on 200 randomly drawn simulations from the 50,547 simulations under the life cycle with host alternation.

Sampling scheme

In the theoretical case presented in the results, we calculate summary statistics based on the maximum information available. We name it the full simulation design, as we consider all time samples, that is populations sampled once a year for 36 years on all host compartments (S and R for the model choice; S, R and A for the parameter estimation). To assess the impact of the sampling scheme on inference accuracy, we consider two types of rarefaction affecting the time series and the range of compartments considered (Appendix A).

2.4 Case study: overcoming of poplar rust resistance RMlp7

Here we apply our ABC inference framework to a documented case of resistance overcoming by a diploid plant pathogen responsible for the poplar rust disease, *Melampsora larici-populina* (Basidiomycota, Pucciniales). This pathogenic fungus is a host-alternating species. Its life cycle consists of an annual sexual reproduction on larch needles in early Spring, followed by clonal multiplications on poplar leaves from spring to autumn (Duplessis et al., 2021; Louet et al., 2021). Its sexual reproduction is obligatory in temperate climates because of the annual fall of poplar leaves (Xhaard, 2011). To control poplar rust disease, several resistant poplars have been selected and planted widely in Western Europe over years. However, the intensive and monocultural plantations combined with the host species being perennial (Gérard et al., 2006) makes poplars particularly vulnerable to the adaptation of the pathogen. This leads

to regular events of resistance overcoming, so that all known resistance types have now been overcome. (Pinon and Frey, 2005; Louet et al., 2021). The most damaging resistance overcoming by *M. larici-populina* occurred in 1994 with the adaptation of the pathogen to the resistance RMlp7 carried at that time by a vast majority of cultivated poplar trees. The adaptation of the pathogen to the resistance RMlp7 resulted from the association of two alterations (a nonsynonymous mutation and a complete deletion) at the candidate locus $AvrMlp7$, from standing genetic variation Louet et al., 2021). This led to a rapid invasion, in less than four years, of adapted pathogens across Western Europe, including France (Barrès et al., 2008; Xhaard et al., 2011; Persoons et al., 2017), causing drastic epidemics (Pinon et al., 1998; Pinon and Frey, 2005). This rapid adaptation event strongly shaped the resulting genetic structure of *M. larici-populina* populations (Xhaard et al., 2011; Persoons et al., 2017).

Poplar rust individuals were sampled before, during, and after this resistance overcoming event and correspond to the samples analysed in the population genetics study of Louet et al., 2021. Each sampled individual was genotyped with 20 microsatellite markers, and population genetics indices are calculated from the model developed above. Two versions of this temporal sampling are used. The first data set is composed of poplar rust samples from the same geographic location (Amance, France), collected from susceptible poplars (S) at the end of the epidemic season and from larch (A) after the sexual reproduction. Twenty-eight populations were sampled between 1989 and 2021 (22 on S, and 6 on A). Sampling size ranges from 5 to 58 individuals (samples with less than 5 individuals were removed from the analyses). The second data set is composed of poplar rust samples from a broader geographic region (Grand Est region, France), collected from susceptible and resistant poplars (S and R) at the end of the epidemic season, and from larch (A) after the sexual reproduction. This second data set includes all individuals from the first data set. Thirty-six populations were sampled between 1988 and 2021 (26 on S, 4 on S and 6 on A). The sampling size ranges from 5 to 79 individuals (samples with less than 5 individuals were removed from the analyses).

For this case study, cross-validation procedures are performed as described above, based on simulated data with the same sampling schemes as the two biological data sets. For the two data sets, the accuracy of the model choice and parameter estimation are compared depending on the choice of summary statistics (Complete summary statistics, or Wrap-up summary statistics). Model choice and parameter estimations are then performed as described above on the empirical data sets, and the posterior distribution is obtained for each model parameter.

For the model choice, we check the goodness-of-fit by computing the $P-value$ to test the fit of the observed data to each model. For each model, we reject the null hypothesis model if $P-value < 0.01$.

3 Results

3.1 Model choice and parameter estimations for the full simulation design

We first evaluate the accuracy of our ABC inference under the ideal case with the maximum data available: populations are sampled every generation from all host compartments.

The cross-validation procedure for the model choice highlights a strong accuracy of model choice when Wrap-up summary statistics are used (Table 3). Conversely, the model identifiability is weaker for Complete summary statistics. However, whatever the summary statistics used, there is a strong overlap in the outcome of simulations realised under the two life cycles (Figure 2). Hence only small areas of parameter values allow to truly discriminate between the two models.

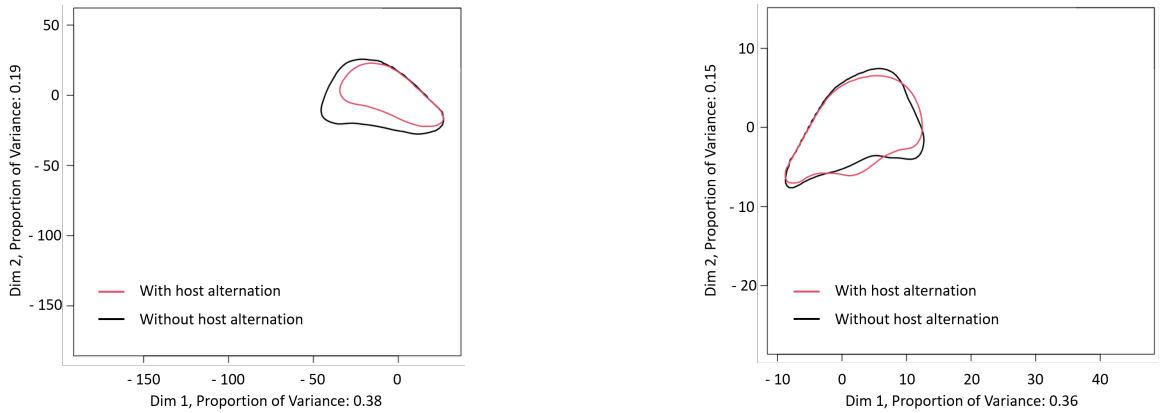
The cross-validation procedure for parameter estimation highlights a strong discrepancy among parameters. Some parameters are very well estimated, including $propR$, K, f_{avr} and r, ranked in a decreasing order of identifiability (Table 4, Figure 3). The worst parameter to estimate is mig, irrespective of the simulated value (Figure 3). Last, τ is badly estimated, but with increasing confidence as its true value increases (Figure 3).

3.2 Model choice and parameter estimations when rarefaction is applied

For both cross-validations, sampling every five years gives similar results than considering the full time series of samplings. Reducing the range of sampled compartments also provides fairly good estimates for the model choice and parameter estimations. This result holds even if the sampling focuses on S only. Overall, the ABC accuracy is higher for Wrap-up summary statistics than for Complete summary statistics. Last, the accuracy of the model choice drops drastically if only the first and last time points are considered (Table S1) and the parameter estimations are less accurate, especially for r and τ. However $propR$, f_{avr} and K are still well estimated, regardless of the rarefaction (Table S2).

Table 3: Accuracy of model selection for the full simulation design depending on the type of summary statistics considered. The model choice procedure is based on leave-one-out cross-validations with a weighted multinomial logistic regression computed with tolerance parameter set at 0.01, for 500 replicates.

Summary statistics	Probability to find the true model	
	Without host alternation	With host alternation
Complete	0.76	0.80
Wrap-up	1.00	0.99



(a) Complete summary statistics

(b) Wrap-up summary statistics

Figure 2: Principal component analyses (95% envelope) of simulations for each model: with and without host alternation, depending on the summary statistics considered. PCA analyses were based on the full simulation design.

Table 4: Accuracy of parameter estimation for the full simulation design depending on the summary statistics considered. Data represent r-squared of the linear regressions between simulated and estimated parameters. The parameters identifiability procedure is based on a leave-one-out cross-validation with the neural networks regression method and tolerance parameter set at 0.01, for 200 replicates.

Summary statistics	$propR$	mig	f_{avr}	r	K	τ
Complete	0.91	0.41	0.68	0.73	0.81	0.58
Wrap-up	0.94	0.60	0.86	0.85	0.92	0.79

3.3 Case study: Approximate Bayesian Computations applied to the poplar rust resistance overcoming

In this section, we apply our ABC framework to the two empirical data sets describing the rapid evolution of the genetic structure of a pathogen population following resistance overcoming.

3.3.1 Cross-validations applied to the sampling schemes of empirical data sets

We first perform the cross-validation procedures with sampling schemes corresponding to the two data sets (Amance location and Grand Est region). For both data sets, the cross-validation of model choice shows weak identifiability of life cycles (Table 5). The identifiability of the life cycle with host alternation is slightly better than without host alternation but still limited. The model is more identifiable with the sampling scheme of the data set from the Grand Est region, which represents more time points and both S and R compartments, contrary to the data set from Amance location sampled on S only. As for the full simulation design, the accuracy of parameter estimations is particularly high for parameters $propR$ and K, regardless

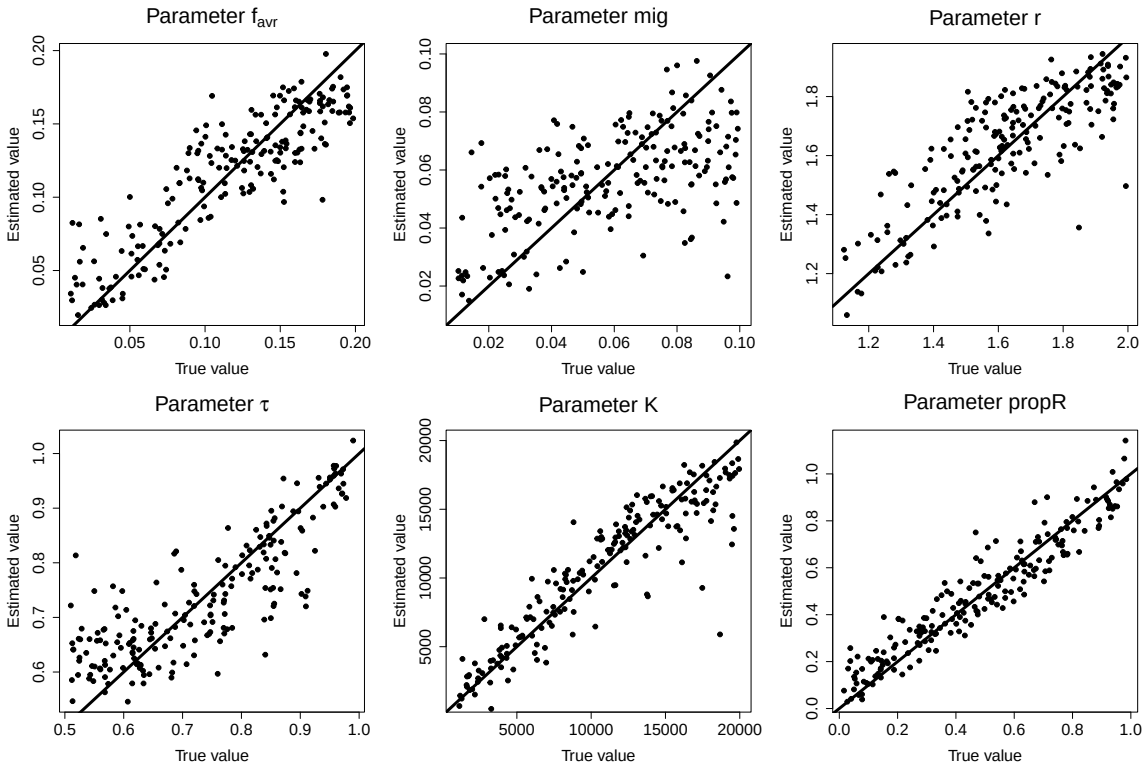


Figure 3: Practical identifiability of parameter estimation for Wrap-up summary statistics. Each point represents the parameter estimation ('Estimated' value) depending on the real parameter ('True' value). Each graph regroups the results of 200 replicates. Straight lines correspond to the first bisector.

of the data set (Table 6). The estimation of parameters f_{avr} is more mitigated but still quite good ($r - squared > 0.7$ for the two data sets), especially with Wrap-up summary statistics. Conversely, mig, r and τ are less accurately estimated, especially with Complete summary statistics. Overall, as for the full simulation design, the ABC performs better using Wrap-up summary statistics than Complete summary statistics.

Table 5: Accuracy of model selection on simulated populations on S and R, depending on the summary statistics considered. The sampling schemes of simulation data sets match those of the two the empirical data sets: Amance location and Grand Est region. The model choice procedure is based on leave-one-out cross-validations with a weighted multinomial logistic regression computed with tolerance parameter set at 0.01, for 500 replicates.

Sampling scheme	Summary statistics	Probability to find the true model	
		Without host alternation	With host alternation
Amance	Complete	0.64	0.78
	Wrap-up	0.67	0.81
Grand Est	Complete	0.73	0.81
	Wrap-up	0.79	0.88

Table 6: Accuracy of the parameter estimation on simulated populations sampled on S, R and A compartments for the life cycle with host alternation, depending on the summary statistics considered. The sampling schemes of the simulation data sets match those of the two the empirical data sets: Amance location and Grand Est region. Numbers represent r-squared of the linear regressions between simulated and estimated parameters. The parameter identifiability procedure is based on a leave-one-out cross-validation with the neural networks regression method and tolerance parameter set at 0.01, for 200 replicates.

Sampling scheme	Summary statistics	$propR$	mig	f_{avr}	r	K	τ
Amance	Complete	0.90	0.27	0.71	0.68	0.83	0.58
	Wrap-up	0.90	0.39	0.77	0.79	0.89	0.67
Grand Est	Complete	0.91	0.31	0.71	0.68	0.81	0.68
	Wrap-up	0.90	0.37	0.77	0.72	0.85	0.67

3.3.2 Model choice

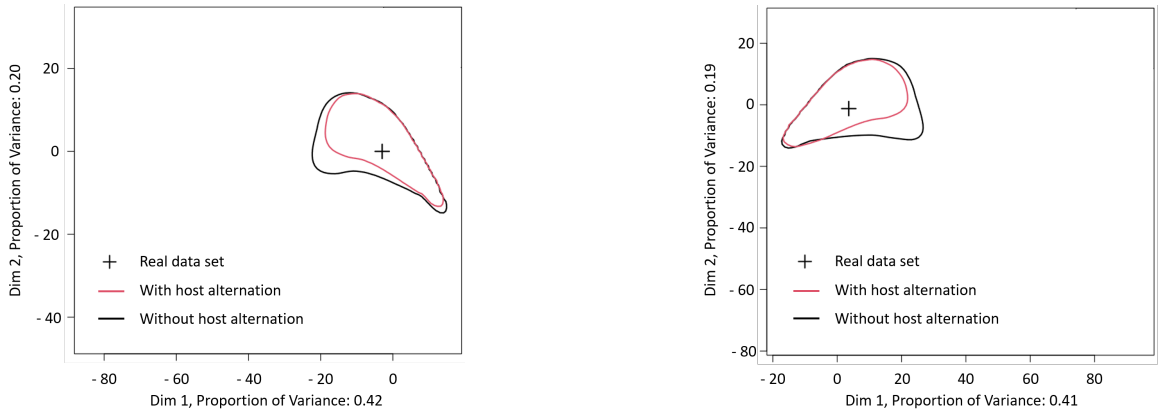
For the two empirical data sets, we infer the life cycle with host alternation with probability 100% from Complete summary statistics and the life cycle without host alternation with probabilities 81% from Wrap-up summary statistics (Table 7). However, in all cases, the results of the goodness-of-fit tests do not allow to significantly reject either of the life cycles, irrespective of the data set considered and the summary statistics used. Therefore, we do not have enough information to properly infer the pathogen life cycle. This is consistent with the coordinates of the data sets on the PCA analyses (Figure 4, which are located in the overlap area of the two models. This explains the weak identifiability of the life cycle and the impossibility of significantly rejecting either life cycle. Therefore, we have to rely on biological insights to determine the life cycle.

Table 7: Posterior model probabilities and goodness of fit for two data sets: in Amance location and the Grand Est region, depending on the summary statistics. With and Without stand for life cycles with and without host alternation, respectively. $P - values < 0.01$ are considered significant.

Sampling scheme	Summary statistics	Posterior probability		Goodness of fit ($P - value$)	
		Without	With	Without	With
Amance	Complete	0.00	1.00	0.069	0.045
	Wrap-up	1.00	0.00	0.048	0.030
Grand Est	Complete	0.00	1.00	0.099	0.062
	Wrap-up	0.81	0.19	0.092	0.064

3.3.3 Parameter estimation

Knowing that the poplar rust pathogen alternates on larch to perform its sexual reproduction, we compute the parameter estimations under the model with host alternation.

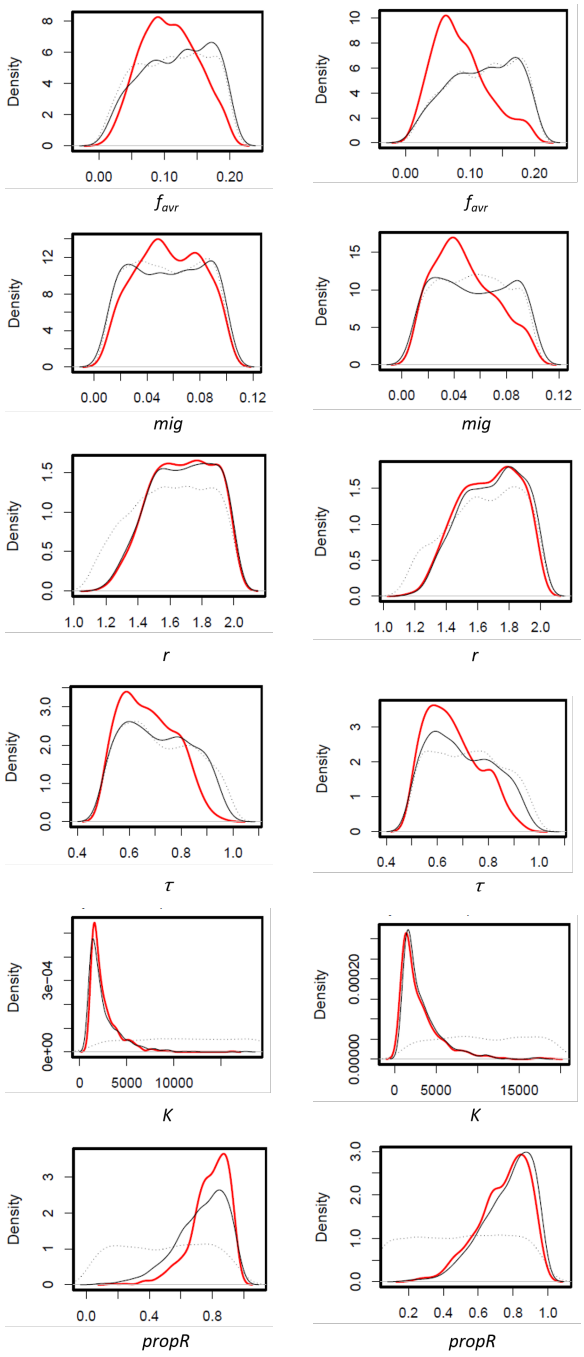


(a) Amance

(b) Grand Est region

Figure 4: Principal component analyses (95% envelope) of simulations for each model: with and without host alternation, depending on the data set considered. Simulations were performed with Complete summary statistics. The black crosses correspond to the coordinates of the actual data sets.

Contrary to the theoretical analysis of parameter identifiability, Complete summary statistics perform better than Wrap-up summary statistics for the analysis of both empirical data sets. For most parameters, especially those that are well identified, the posterior distribution is more narrow and the difference between the posterior density versus the prior density is more pronounced (Figures 5, S1). Therefore, in the following, we focus on the parameter estimation from Complete summary statistics. As expected from the theoretical identifiability, narrow posterior distributions allow a confident inference of parameters $propR$ and K. The strength of inference of parameters f_{avr} and τ is more limited, and the posteriors for parameters mig and r are not much informative compared to their prior distributions. From the mode of each posterior distribution, we infer a high proportion of resistant poplars at the time of resistance overcoming: $propR = 0.78$ and $propR = 0.84$ for Amance and the Grand Est region data sets respectively. The inferred population sizes are in the order of a thousand (mode values $K = 1,523$ and $1,356$ for Amance and the Grand Est region, respectively). We infer an initial proportion of virulent alleles f_{avr} in the pathogen population of 9% in Amance, and 6% in the Grand Est region, but with a large credible interval (range of values for which the posterior distribution is higher than the prior distribution), between 5% and 15%. The annual mortality rate τ during the annual migration preceding sexual reproduction is inferred to close values of 0.60 and 0.58 in Amance and in the Grand region, respectively. The migration rate mig and growth rate r are badly estimated (mig would range from 0.02 to 0.10, and r would be above 1.5).



(a) Amance (b) Grand Est region

Figure 5: Posterior distributions of parameters, with Complete summary statistics, for the data sets in Amance location (on S and A) and the Grand Est region (on S, R and A). Dashed lines correspond to the prior distributions, black lines correspond to posterior distributions given by the rejection method, red lines correspond to the posterior distributions given by the neuralnet method.

Table 8: Statistical summary of the inference of the parameters for the life cycle with host alternation, depending on the data set.

Data set	*Parameter*	$q-2.5\%$	*median*	*mean*	*mode*	$q-97.5\%$
Amance	$propR$	0.49	0.81	0.78	**0.88**	0.94
	mig	0.01	0.06	0.06	**0.05**	0.10
	f_{avr}	0.03	0.11	0.11	**0.09**	0.19
	r	1.31	1.69	1.68	**1.92**	1.98
	K	1,205	2,024	2,561	**1,374**	6,414
	τ	0.51	0.66	0.67	**0.58**	0.88
Grand Est	$propR$	0.42	0.77	0.74	**0.80**	0.94
	mig	0.01	0.04	0.05	**0.04**	0.10
	f_{avr}	0.02	0.08	0.09	**0.06**	0.19
	r	1.31	1.69	1.67	**1.80**	1.96
	K	768	2,206	3,108	**1,273**	10,146
	τ	0.51	0.64	0.66	**0.58**	0.88

4 Discussion

Time samples unravel rapid adaptation

In this paper, we develop an original methodology to infer demographic scenarios and parameter values in the case of rapid adaptation. We base our inference framework on the use of time series data to grasp changes in population structure over time. We employ a forward population genetics model to simulate the summary statistics used for ABC inference. As such, our methodology contrasts with the most recent developments in population genetics inference that are based on the coalescent theory. Theoretical investigations demonstrate high accuracy in both model and parameter estimations. Last, our inferential framework is successfully applied to a case study of rapid adaptation in a plant pathogen following a resistance overcoming event with sound estimates of key demographic parameters.

Too many summary statistics lead to model overfitting

The main originality of our method is to take into account the time series, compared to the high amount of studies that focus on a single time point. To do so, we adapt the classical ABC protocols of using summary statistics. We compare two methods, either keeping the complete sequence of statistics over all time samples or wrapping up the information into the mean, variance, minimum and maximum values for each summary statistic. We show in the theoretical cross-validation procedures that Wrap-up summary statistics outperform Complete summary statistics. This result may appear counter-intuitive. However too much and redundant information can lead to a decrease in statistical power. Since many statistics can be used in ABC, there is a certain curse of dimensionality, early on identified. This generates overfitting that may affect the accuracy of model and parameter estimation. This is especially acute

when several summary statistics correlate with the same model parameters, inflating the heterogeneous accuracy inference of some parameters compared to others. A suggested solution is to perform first a dimension reduction of the statistics space before performing the ABC estimations. In our case, we have few genetic indices and few parameters, so we do not perform such reduction of dimensionality beforehand. We want to keep the full time series to avoid loss of temporal information, but in doing so we inflate the number of summary statistics which is likely to result in temporal auto-correlation causing overfitting. In support of this hypothesis, the discrepancy in parameter identifiability between Complete and Wrap-up statistics tends to decrease when we apply a thinning in the time series consisting in keeping only one fifth of the temporal samples. A way to keep the maximum information while avoiding the redundancy that causes overfitting would be to summarise the time series in a different manner. For example, we may fit a function describing the temporal changes and use regression coefficients as summary statistics. However, this method requires that all dynamics be correctly fitted by a similar function (with the same number of regression coefficients). In this attempt, we try to fit quadratic polynomes, but the shape of the temporal variation varied too much among population genetic indices to provide conclusive summary statistics (data not shown). Moreover, such fit cannot take into account stochastic variation due to drift. Another way to improve the inferential framework would be to avoid the use of population genetics indices as summary statistics and base the inference method directly on the evolution of genotypic frequencies, or a less condensed type of information such as site frequency spectra (SFS) across time samples. However this method may be more computationally intensive (in particular for storing the results over a long time series), and the SFS cannot be computed from microsatellite markers. Moreover, the resulting increase in the number of summary statistics could also lead to high dimensionality of the data.

Case study estimates match the biology of the poplar rust system

By applying this method on empirical data sets, we infer accurately three parameters: $propR$, K and f_{avr}. The estimated values make sense with regard to our knowledge of this event of resistance overcoming. We infer a very large proportion of resistant poplars (80%), which is consistent with our knowledge of poplar plantations before the RMlp7 resistance overcoming. Indeed, the resistant poplar cultivar 'Beaupré' that bears resistance RMlp7 was widely planted at the time of resistance overcoming and represented up to 80% of poplar cultivar sales in 1996 (data from the French Ministry of Agriculture, Fabre et al., 2021). This very high proportion of resistant poplars in the landscape exerted a strong selection pressure and accounts for the changes in the genetic structure of a pathogen population over time (Persoons et al., 2017).

We infer a population size of around one thousand individuals. This represents the effective population size rather than the population size, *i.e.* the number of individuals that effectively contribute to the observed genetic variability. This order of magnitude is consistent with a previous estimate based on a coalescent analysis (Persoons, 2015). We estimate an initial proportion of virulent alleles in the population between 5% and 10%. This range is consistent with the genetic characterisation at the selected locus in *M. larici populina* (Louet et al., 2021). Louet et al. (2021) highlighted that the alleles conferring virulence pre-existed in the pathogen population long before the resistance overcoming, at a frequency of 21%, on average, between 1989 and 1993. The proportion of virulent alleles in the population can strongly fluctuate during the years preceding resistance overcoming (Saubin et al., 2021). Such fluctuations are indeed observed in the data from Louet et al. (2021), with a standard deviation in the proportion of virulent alleles of 0.12 between 1989 and 1993. Our estimation is therefore consistent with the empirical data considering such large fluctuations from year to year. This level of standing genetic variation is not negligible in view of the adaptive potential it brings to the pathogen population.

The fact that some parameters are not well estimated is not caused by a limited amount of data, but rather by a strong assumption in our modelling framework. Indeed, we assume three compartments only in the landscape and neglect the more complex spatial structures that can be encountered in agricultural landscapes. In such a non-spatial system, it may be especially difficult to disentangle the genetic effects of growth, mortality and migration rates. We believe that increasing the spatial complexity of the model would help disentangle these three parameters. However, this would also highly increase the dimensions of summary statistics, which can lead to a decrease in the statistical accuracy of the ABC.

Despite the precautions taken when evaluating the theoretical accuracy of our method, we observe that the accuracy of our ABC estimation differs between the theoretical and empirical results. We obtain better theoretical results with Wrap-up summary statistics than with Complete summary statistics. However, Complete summary statistics lead to more accurate parameter inferences from empirical data sets. We believe that this is due to increased variability in population genetics indices in the empirical data compared to the simulated data. Many sources of biological stochasticity are not accounted for in our modelling assumptions. This stochasticity results in stronger local extremes, which are captured by the minimum and maximum values in Wrap-up summary statistics. Thus, taking into account all the available information of the empirical time series through Complete summary statistics can reduce the impact of this additional stochasticity, and improve the parameter estimations. This discrepancy between our application to empirical data and theoretical findings from simulated data

can also originate from strong modelling assumptions (like the non-spatial model) that do not perfectly reflect the biological system.

Methodological guideline

We recommend using a sampling scheme with regular time points to capture sufficient information in the genetic evolution of populations. Contrary to the current practice, we show that for this framework it is more efficient to sample more populations in time but on a single host than to focus on few temporal samples but on several hosts. The sampling scheme with the first and the last time points, even if less informative than a more regular sampling, still allows to correctly infer some of the model parameters. In particular, the initial frequency of virulent alleles in the population is very well inferred with this sampling scheme because the first population in time is the most important to identify the initial genetic composition of the population.

Future directions: application for genome scans of rapid adaptation

Neutral genetic data from microsatellite markers contain sufficient information to infer parameters of resistance overcoming from time samples. However, we could still gain in accuracy by taking genomic data into account. The integration of genomic data would allow the calculation of a wider range of summary statistics, which could increase the accuracy of inferences. Fitting demographic models is a prerequisite for detecting selection, but it can be difficult to do in practice and is often inaccurate (Hoban et al., 2016). The addition of genomic data to the described framework would enable to focus on areas under selection, detect sweeps (Messer and Petrov, 2013; Foll et al., 2015) and calculate their age. Determining the evolution of neutral loci through such a selection event may allow, by comparison, to identify selected loci implied in rapid adaptation.

Acknowledgements

We warmly thank Emma Chavan for her help with the microsatellite analyses. This work was supported by grants from the French National Research Agency (ANR-18-CE32-0001, CLONIX2D project; ANR-11-LABX-0002-01, Cluster of Excellence ARBRE) and from the Metaprogram SUMCROP of the National Research Institute for Agriculture, Food and the Environment (INRAE, Opiniâtres project). Méline Saubin was supported by a PhD fellowship from INRAE and the French National Research Agency (ANR-18-CE32-0001, CLONIX2D project). Méline Saubin obtained an international mobility grant BAYFRANCE as part of a Franco-

Bavarian cooperation project, to work for one month in Aurélien Tellier's lab (Grant Number FK21_2020).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agapow, P. M. and Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, 1(1-2):101–102.

Agrios, G. N. (2005). *Plant pathology*. Elsevier edition.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., and Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution*, 19(10):536–544.

Arnaud-Haond, S., Duarte, C. M., Alberto, F., and Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, 16:5115–5139.

Barrès, B., Halkett, F., Dutech, C., Andrieux, A., Pinon, J., and Frey, P. (2008). Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, 8(5):577–587.

Bazin, E., Dawson, K. J., and Beaumont, M. A. (2010). Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics*, 185(2):587–602.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035.

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., and Petrov, D. A. (2014). Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLoS Genetics*, 10(11).

Blum, M. G. B. and François, O. (2010). Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, 20(1):63–73.

Bonneaud, C. and Longdon, B. (2020). Using evolutionary theory to understand the fate of novel infectious pathogens. *Science and Society*, 21(e51374).

Brown, J. K. M. and Tellier, A. (2011). Plant-parasite coevolution: Bridging the gap between genetics and ecology. *Annual Review of Phytopathology*, 49(1):345–367.

Buffalo, V. and Coop, G. (2019). *The linked selection signature of rapid adaptation in temporal genomic data*, volume 213.

Buffalo, V. and Coop, G. (2020). Estimating the genome-wide contribution of selection to temporal allele frequency change. *Proceedings of the National Academy of Sciences of the United States of America*, 117(34):20672–20680.

Burdon, J. J., Zhan, J., Barrett, L. G., Papaix, J., and Thrall, P. H. (2016). Addressing the challenges of pathogen evolution on the world's arable crops. *Phytopathology*, 106(10):1117–1127.

Cornuet, J. M., Ravigné, V., and Estoup, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, 11(401).

Csilléry, K., François, O., and Blum, M. G. B. (2012). Abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3(3):475–479.

De Mita, S. and Siol, M. (2012). EggLib: Processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics*, 13(1):27.

Dehasque, M., Ávila-Arcos, M. C., Díez-del Molino, D., Fumagalli, M., Guschanski, K., Lorenzen, E. D., Malaspinas, A. S., Marques-Bonet, T., Martin, M. D., Murray, G. G. R., Papadopulos, A. S. T., Therkildsen, N. O., Wegmann, D., Dalén, L., and Foote, A. D. (2020). Inference of natural selection from ancient DNA. *Evolution Letters*, 4(2):94–108.

Dobzhansky, H. (1943). Genetics of natural populations IX. Temporal changes in the composition of populations of *Drosophila pseudoobscura*. *Genetics*, 28:162–186.

Duplessis, S., Lorrain, C., Petre, B., Figueroa, M., Dodds, P. N., and Aime, M. C. (2021). Host adaptation and virulence in heteroecious rust fungi. *Annual Review of Phytopathology*, 59:403–422.

Estoup, A. and Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: Why, how and so what? *Molecular Ecology*, 19(19):4113–4130.

Fabre, B., Bastien, C., Husson, C., Marçais, B., Frey, P., and Halkett, F. (2021). Chapitre 27. Un effet papillon dans les peupleraies françaises : les répercussions d'un contournement de résistance sur les méthodes de sélection variétale. In Lannou, C., Roby, D., Ravigné, V., Hannachi, M., and Moury, B., editors, *L'immunité des plantes*, pages 329–339. Versailles, quae edition.

Fagundes, N. J., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F. M., Bonatto, S. L., and Excoffier, L. (2007). Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 104(45):17614–17619.

Feder, A. F., Pennings, P. S., and Petrov, D. A. (2021). The clarifying role of time series data in the population genetics of HIV. *PLoS Genetics*, 17(1):1–10.

Fisher, R. A. and Ford, E. B. (1947). The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity*, 1(2):143–174.

Foll, M., Shim, H., and Jensen, J. D. (2015). WFABC: A Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources*, 15(1):87–98.

Gérard, P. R., Husson, C., Pinon, J., and Frey, P. (2006). Comparison of genetic and virulence diversity of *Melampsora larici-populina* populations on wild and cultivated poplar and influence of the alternate host. *Phytopathology*, 96(9):1027–1036.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Hartfield, M. (2021). Approximating the Coalescent Under Facultative Sex. *The Journal of heredity*, 112(1):145–154.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., and Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4):379–397.

Johnson, R. (1984). A critical analysis of durable resistance. *Annual Review of Phytopathology*, 22.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kettlewell, H. B. D. (1958). A survey of the frequencies of *Biston betularia* (L.) (Lep.) and its melanic forms in great britain. *Heredity (Edinb.)*, 12:51–72.

Kettlewell, H. B. D. (1961). The phenomenon of industrial melanism in Lepidoptera. *Annual Review of Entomology*, 6(1):245–262.

Laval, G., Patin, E., Boutillier, P., and Quintant-Murci, L. (2019). A genome-wide Approximate Bayesian Computation approach suggests only limited numbers of soft sweeps in humans over the last 100,000 years. *bioRxiv*, pages 1–53.

Louet, C., Saubin, M., Andrieux, A., Persoons, A., Gorse, M., Pétrowski, J., Fabre, B., De Mita, S., Duplessis, S., Frey, P., and Halkett, F. (2021). A point mutation and large deletion at the candidate avirulence locus *AvrMlp7* in the poplar rust fungus correlate with poplar RMlp7 resistance breakdown. *Molecular Ecology*, (February):1–12.

McDonald, B. A. and Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, 40:349–379.

Messer, P. W. and Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, 28(11):659–669.

Mueller, L. D., Barr, L. G., and Ayala, F. J. (1985a). Natural selection vs. random drift: Evidence from temporal variation in allele frequencies in nature. *Genetics*, 111(3):517–554.

Mueller, L. D., Wilcox, B. A., Ehrlich, P. R., Heckel, D. G., and Murphy, D. D. (1985b). A direct assessment of the role of genetic drift in determining allele frequency variation in populations of *Euphydryas editha*. *Genetics*, 110(3):495–511.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3):583–590.

Nei, M. and Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics*, 98(3):625–640.

Nielsen, R. and Signorovitch, J. (2003). Correcting for ascertainment biases when analyzing SNP data: Applications to the estimation of linkage disequilibrium. *Theoretical Population Biology*, 63(3):245–255.

Orive, M. E. (1993). Effective population size in organisms with complex life-histories.

Parat, F., Schwertfirm, G., Rudolph, U., Miedaner, T., Korzun, V., Bauer, E., Schön, C. C., and Tellier, A. (2016). Geography and end use drive the diversification of worldwide winter rye populations. *Molecular Ecology*, 25(2):500–514.

Parsons, T. L., Lambert, A., Day, T., and Gandon, S. (2018). Pathogen evolution in finite populations : slow and steady spreads the best. *Journal of royal society*, 15.

Pavinato, V. A. C., De Mita, S., Marin, J. M., and de Navascués, M. (2021). Joint inference of adaptive and demographic history from temporal population genomic data. *bioRxiv*, pages 1–18.

Persoons, A. (2015). *Les contournements de résistance par Melampsora larici-populina, l'agent de la rouille du peuplier : impact démographique et déterminisme génétique*. PhD thesis, Université de Lorraine.

Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., and Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, 26(7):1902–1918.

Pink, D. and Puddephat, I. (1999). Deployment of disease resistance genes by plant transformation - a 'mix and match' approach. *Trends in Plant Science*, 4(2):71–75.

Pinon, J. and Frey, P. (2005). Interactions between poplar clone and *Melampsora* populations and their implications for breeding for durable resistance. *Rust Diseases of Willow and Poplar*, (July):138–154.

Pinon, J., Frey, P., Husson, C., and Schipfer, A. (1998). Poplar rust (*Melampsora larici-populina*): the development of E4 pathotypes in France since 1994. *First IUFRO Rusts of Forest Trees*, 712:57–64.

Pollak, E. (1983). A new method for estimating the effective population size from allele frequency changes. *Genetics*, 104(3):531–548.

Prout, T. (1954). Genetic drift in irradiated experimental populations of *Drosophila melanogaster*. *Genetics*, 39(4):529–545.

Rimbaud, L., Fabre, F., Papaïx, J., Moury, B., Lannou, C., Barrett, L. G., and Thrall, P. H. (2021). Models of plant resistance deployment. *Annual Review of Phytopathology*, 59(1).

Rosenberg, N. A. and Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, 3(5):380–390.

Saubin, M., De Mita, S., Zhu, X., Sudret, B., and Halkett, F. (2021). Impact of ploidy and pathogen life cycle on resistance durability. *Peer Community Journal*, 1(e8):1–31.

Saubin, M., Louet, C., Bousset, L., Fabre, F., Frey, P., Fudal, I., Grognard, F., Hamelin, F., Mailleret, L., Stoeckel, S., Touzeau, S., Petre, B., and Halkett, F. (2022). Improving sustainable crop protection using population genetics concepts. *Molecular Ecology*, 00:1–11.

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., Mcroberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature ecology and evolution*, 3(3):430–439.

Sjödin, P., Kaj, I., Krone, S., Lascoux, M., and Nordborg, M. (2005). On the meaning and existence of an effective population size. *Genetics*, 169(2):1061–1070.

Stukenbrock, E. H. and McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, 46:75–100.

Tellier, A. and Lemaire, C. (2014). Coalescence 2.0: A multiple branching of recent theoretical developments and their applications. *Molecular Ecology*, 23(11):2637–2652.

Tobin, P. C. (2015). Ecological consequences of pathogen and insect invasions. *Forest pathology and entomology*, 1:25–32.

van Rossum, G. (1995). Python tutorial, Technical Report CS-R9526. *CWI*.

Wallace, B. (1956). Studies on irradiated populations of *Drosophila melanogaster*. *Journal of Genetics*, 54(2):280–293.

Wang, J. and Whitlock, M. C. (2003). Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, 163(1):429–446.

Waples, R. S. (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, 121(2):379–391.

Whitlock, M. C. and Barton, N. H. (1997). The effective size of subdivided population. *Genetics*, 146(1):427–441.

Wright, S. (1949). *Adaptation and selection*, chapter Adaptation, pages 365–389. Princeton Univ. Press, Princeton, NJ.

Wright, S. (1978). *Evolution and the genetics of populations - Variability within and among natural populations*. The University of Chicago Press.

Xhaard, C. (2011). *Influence des processus démographiques sur la structure et les caractéristiques génétiques des champignons phytopathogènes, cas de l'agent de la rouille du peuplier Melampsora larici-populina*. PhD thesis.

Xhaard, C., Fabre, B., Andrieux, A., Gladieux, P., Barrès, B., Frey, P., and Halkett, F. (2011). The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. *Molecular Ecology*, 20(13):2739–2755.

Zhan, J., Thrall, P. H., Papaïx, J., Xie, L., and Burdon, J. J. (2015). Playing on a pathogen's weakness: Using evolution to guide sustainable plant disease control strategies. *Annual Review of Phytopathology*, 53(1):19–43.

Živković, D., John, S., Verin, M., Stephan, W., and Tellier, A. (2019). Neutral genomic signatures of host-parasite coevolution. *BMC Evolutionary Biology*, 19(1):1–11.

Appendix

A Effect of sampling rarefaction applied to compartments and time-series.

To investigate the effect of a reduced amount of information, we perform ABC analyses based on rarefied sampling scheme. Cross-validation procedures for the model choice (Table S1) and parameter estimation (Table S2) are performed as presented in Section 2.3.

Two types of rarefaction are applied, affecting the time series and the range of compartments considered.

Concerning the time series, in addition to the full time series, we consider (1) a thinning that keeps samplings every five years for a total of eight time samples and (2) only the first and last time samples. For the latter temporal rarefaction, it is not possible to compare the two types of summary statistic since only two generations are considered.

Concerning rarefaction of the sampled compartment (S, R or A) we apply different rarefaction types depending on the cross-validation procedure. The model choice procedure is based on: (1) Populations sampled on both S and R; (2) Populations sampled on S only. Populations sampled on A are not used for the model choice because this compartment only exists for the life cycle with host alternation. The parameter estimation procedure focus on the life cycle with host alternation and is based on: (1) Populations on S, R, and A; (2) Populations on S and R; (3) Populations on S and A; (4) Populations on S only.

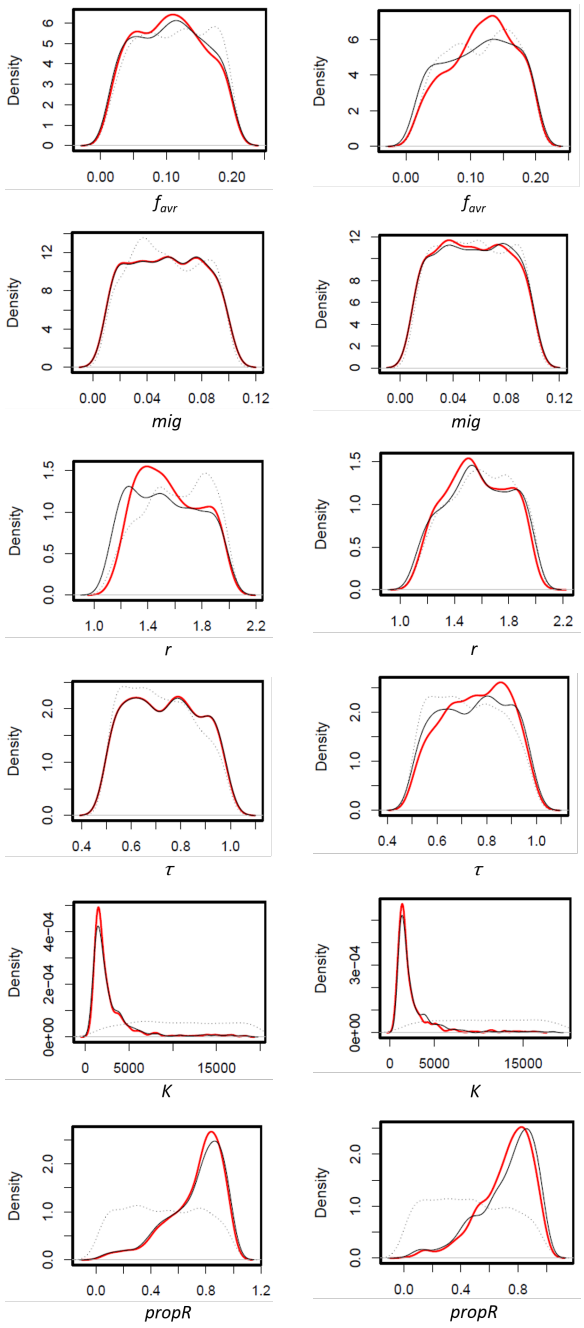
Table S1: Accuracy of model selection depending on the summary statistics and the sampling rarefaction considered. The model choice procedure is based on leave-one-out cross-validations with a weighted multinomial logistic regression computed with tolerance parameter set at 0.01, for 500 replicates. Bold values represent the values reported in the main text.

Time samples	Compartment samples	Summary statistics	Probability to find the true model	
			Without alternation	With alternation
Every year	S and R	Complete	**0.76**	**0.80**
		Wrap-up	**1.00**	**0.99**
	S	Complete	0.59	0.79
		Wrap-up	0.97	0.94
Every five year	S and R	Complete	0.76	0.80
		Wrap-up	1.00	0.99
	S	Complete	0.60	0.70
		Wrap-up	0.98	0.97
First and last year	S and R	-	0.57	0.63
	S	-	0.54	0.62

Table S2: Accuracy of the parameter estimation depending on the summary statistics considered and the type of rarefaction, for simulated populations with host alternation. Data represent r-squared of the linear regressions between simulated and estimated parameters. The parameters identifiability procedure is based on a leave-one-out cross-validation with the neural networks regression method and tolerance parameter set at 0.01, for 200 replicates. Bold values represent the values already presented in the main text.

Time samples	Compartment samples	Summary statistics	$propR$	mig	f_{avr}	r	K	τ
Every year	S, R and A	Complete	**0.91**	**0.41**	**0.68**	**0.73**	**0.81**	**0.58**
		Wrap-up	**0.94**	**0.60**	**0.86**	**0.85**	**0.92**	**0.79**
	S and R	Complete	0.80	0.33	0.75	0.78	0.79	0.56
		Wrap-up	0.94	0.70	0.85	0.83	0.90	0.77
	S and A	Complete	0.90	0.23	0.67	0.65	0.79	0.62
		Wrap-up	0.94	0.43	0.83	0.79	0.90	0.78
	S	Complete	0.89	0.31	0.70	0.63	0.85	0.64
		Wrap-up	0.91	0.51	0.81	0.73	0.88	0.73
Every five year	S, R and A	Complete	0.89	0.31	0.73	0.68	0.80	0.68
		Wrap-up	0.90	0.46	0.82	0.68	0.85	0.69
	S and R	Complete	0.88	0.33	0.78	0.60	0.82	0.65
		Wrap-up	0.92	0.31	0.85	0.68	0.86	0.69
	S and A	Complete	0.91	0.32	0.78	0.67	0.80	0.72
		Wrap-up	0.90	0.27	0.79	0.65	0.84	0.69
	S	Complete	0.87	0.23	0.76	0.62	0.82	0.63
		Wrap-up	0.89	0.46	0.81	0.61	0.85	0.65
First and last year	S, R and A	-	0.87	0.45	0.85	0.52	0.83	0.64
	S and R	-	0.81	0.46	0.83	0.30	0.81	0.58
	S and A	-	0.78	0.26	0.83	0.54	0.82	0.68
	S	-	0.82	0.24	0.74	0.34	0.79	0.51

B Case study: parameters' inference with Wrap-up summary statistics



(a) Amance (b) Grand Est region

Figure S1: Posterior distributions of parameters, with Wrap-up summary statistics, for the datasets in Amance location (on S and A) and the Grand Est region (on S, R and A). Dashed lines correspond to the prior distribution, black lines correspond to posterior distributions given by the rejection method, red lines correspond to the posterior distributions given by the neuralnet method.

4.2 Inférence du scénario de contournement de la résistance Rmlp7

À travers cette analyse, nous avons montré qu'une méthode d'inférence par analyse ABC est plus précise avec un échantillonage temporel régulier qu'avec uniquement le premier et le dernier point dans le temps. Il est également plus avantageux d'échantillonner d'avantage d'individus dans le temps mais sur un seul hôte, que de se concentrer sur peu d'échantillons temporels mais sur plusieurs hôtes.

La méthode d'inférence par analyse ABC développée dans ce chapitre nous permet d'estimer avec précision trois paramètres écologiques et démographiques du contournement de la résistance RMlp7 : la proportion de peupliers résistants RMlp7, la taille maximum de la population de Mlp et la proportion initiale des allèles virulents dans la population de Mlp. Les valeurs estimées et leurs ordres de grandeur font sens en regard de nos connaissances du système biologique. La proportion de peupliers résistants est estimée à plus de 80% dans le paysage agricole. Cette valeur est en accord avec notre connaissance des compositions des peupleraies cultivées lors du contournement (FABRE *et al.*, 2021a). La taille de la population de Mlp est estimée à plus d'un millier d'individus. Cette taille de population ne représente pas le nombre d'individus se multipliant réellement dans la population, mais ceux qui contribuent effectivement à la variabilité génétique observée. Cet ordre de grandeur est cohérent avec une estimation précédente basée sur une analyse coalescente (PERSOONS, 2015). La proportion initiale d'allèles virulents dans la population est estimée entre 5% et 10%. Cet ordre de grandeur est cohérent avec les données dont nous disposons suite à la caractérisation génétique au locus sélectionné (LOUET *et al.*, 2021). Parmi les individus échantillonnées entre 1989 et 1993, les allèles virulents avr et Ø représentent 21% des allèles, et sont systématiquement présents à l'état hétérozygote avec un allèle avirulent Avr, et donc au sein d'individus phénotypiquement avirulents. La proportion d'allèles virulents dans la population peut fortement fluctuer au cours des années précédant le contournement (Chapitre 2, Article 2). De telles fluctuations de la proportion d'allèles virulents sont effectivement observées dans les données de LOUET *et al.* (2021), avec un écart type de 0,12 entre 1989 et 1993. D'après les signatures génétiques résultant de ce contournement, nous inférons donc une proportion d'allèles virulents légèrement inférieure à celle retrouvée dans les données empiriques, mais cohérente avec ces données empiriques compte tenu de fluctuations aussi importantes d'une année sur l'autre. Cette forte variabilité génétique préexistante a considérablement augmenté le potentiel adaptatif de Mlp envers les peupliers résistants Rmlp7.

À l'inverse, trois paramètres ne sont pas bien estimés par notre méthode d'inférence : le taux de migration, le taux de croissance et le taux de mortalité de la population de Mlp. L'estimation du taux de migration est trop imprécise pour être interprétée, tandis que les

estimations du taux de croissance et du taux de mortalité, bien qu'imprécises, permettent tout de même d'établir des intervalles de valeurs probables. Le taux de croissance peut vraisemblablement être estimé entre 1.5 et 1.9, ce qui représente un taux élevé. Cette valeur élevée peut expliquer pourquoi il n'y a pas de signal génétique plus prononcé, qui aurait témoigné d'un évènement de sauvetage évolutif comme discuté dans le Chapitre 2 et le Chapitre 3. Le taux de mortalité est estimé entre 0.5 et 0.8 lors de la migration annuelle des individus sur les aiguilles de Mélèze. Ce taux de mortalité peut paraître faible pour des spores se dispersant aléatoirement par le vent. Cependant, en considérant que nous modélisons uniquement les individus qui contribuent effectivement à la variabilité génétique observée, cet ordre de grandeur correspond aux attendus biologiques.

Notre analyse montre que le manque de précision pour l'estimation de ces trois derniers paramètres n'est pas dû à la quantité d'individus échantillonnés ou à la fréquence des échantillonnages, mais plutôt à une forte hypothèse de modélisation. Nous négligeons dans ce modèle la complexité des structures spatiales des paysages agricoles. Dans un tel système non spatialisé, il peut être particulièrement difficile de distinguer les effets génétiques de paramètres antagonistes tels que les taux de croissance, de mortalité et de migration.

Conclusion de la partie I

La première partie de cette thèse a permis de mettre en évidence le lien entre la démographie et l'évolution génétique d'une population d'agents pathogènes responsables d'une maladie émergente lors d'un contournement de résistance. J'ai ainsi déterminé les conditions d'existence d'un contournement de résistance, identifié les déterminants de la durabilité d'une résistance, établi une typologie des scénarios évolutifs et démographiques des agents pathogènes lors d'un contournement, et appliqué ce cadre d'analyse démogénétique théorique à des données empiriques pour inférer les paramètres régissant le contournement de la résistance RMlp7. Ce système écologique de contournement de résistance, où la composante temporelle est déterminante, nous a ainsi permis d'étudier l'évolution de populations d'agents pathogènes lorsqu'elles présentent de fortes variations démographiques. La non spatialisation du modèle représente cependant une hypothèse très forte et a montré ses limites à plusieurs reprises, notamment lors de l'application du modèle aux données au cours de l'inférence. L'expansion géographique d'un agent pathogène responsable d'une maladie émergente nécessite d'être prise en compte pour étudier plus finement le lien entre la démographie et l'évolution génétique d'une telle population.

Deuxième partie

ÉTUDE D'UNE MALADIE RÉ-ÉMERGENTE EN SYSTÈME SPATIALISÉ : LE CAS D'UNE COLONISATION CLONALE

Objectifs

1. Inférer les paramètres de la dynamique d'expansion régissant la colonisation de Mlp dans la vallée de la Durance.
2. Déterminer la variabilité des structures génétiques observées empiriquement au cours de différents évènements de colonisation de Mlp dans la vallée de la Durance.
3. Déterminer l'influence de la dynamique d'expansion de l'agent pathogène sur les structures génétiques résultantes.

Chapitre 5 Introduction **199**
5.1 Dispersion et démographie . . . 201
5.2 Structures génétiques sur un front de colonisation . . . 203
5.3 Cas d'étude : les invasions récurrentes de Mlp le long de la vallée de la Durance 206
5.4 Objectifs . . . 207

Chapitre 6 Modèle mécanico-statistique et inférence de la dynamique d'expansion de Mlp dans la vallée de la Durance **209**
6.1 Article 5. Inferring invasion determinants with mechanistic models and multitype samples . . . 211
6.2 Dynamique d'expansion de Mlp dans la vallée de la Durance . . . 246

Chapitre 7 Variabilité inter-annuelle de la structure génétique de Mlp dans la vallée de la Durance **249**
7.1 Article 6. Genetic signatures of a range expansion *in natura*: variation across years . . . 251
7.2 Dynamique d'expansion et structures génétiques . . . 272

Conclusion de la partie II **273**

Chapitre 5

Introduction

5.1 Dispersion et démographie

La dynamique des populations le long d'un front de colonisation résulte de quatre principaux mécanismes écologiques (MILLER *et al.*, 2020, FIGURE 5.1) :

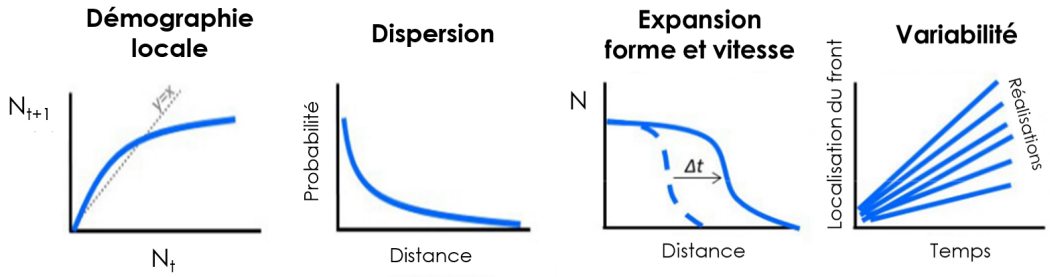


FIGURE 5.1 – Schéma des quatre principaux mécanismes écologiques impactant la dynamique d'une population sur un front de colonisation (adapté d'après MILLER *et al.*, 2020).

- **La démographie** représente ici l'ensemble des évènements régissant le succès reproductif d'un individu au cours de sa vie. Sur un front de colonisation, ce succès reproductif peut varier localement selon la taille de la population et il diminue lorsque cette taille se rapproche de la capacité de charge de l'environnement.
- **La dispersion** représente la redistribution des individus (généralement une descendance) depuis un point source. Différentes fonctions peuvent être utilisées pour décrire la dispersion des individus sur un front de colonisation (Encadré 5.1). Ces fonctions permettent de prendre en compte plus ou moins d'évènements de dispersion à longue distance (LDD).
- **La forme et la vitesse de l'expansion** caractérisent l'évolution temporelle de la distribution de la densité en individus le long d'un front de colonisation. Elles sont très dépendantes de la fonction de dispersion.
- **La variabilité de l'expansion** correspond à l'hétérogénéité des réalisations d'un processus d'expansion géographique pour des conditions initiales identiques.

Parmi ces mécanismes, deux sont étroitement liés et interviennent à l'échelle de l'individu lors d'une expansion géographique : la démographie et la dispersion (MILLER *et al.*, 2020). Les régimes de dispersion peuvent fortement impacter la dynamique des populations. Les évènements de LDD fréquents peuvent participer, par exemple, à l'accélération d'un front de colonisation (MEDLOCK ET KOT, 2003; GARNIER, 2011; BOUIN *et al.*, 2018).

Encadré 5.1 - Les fonctions de dispersion

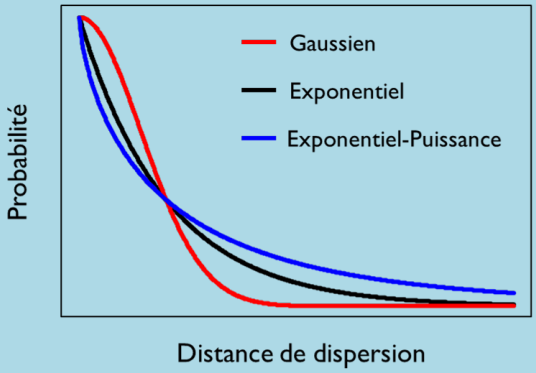


FIGURE 5.2 – Fonctions de densité de trois noyaux de dispersion menant à différentes fréquences d'évènements de dispersion à longue distance.

La diffusion est régulièrement utilisée pour modéliser les déplacement des individus de proche en proche lors d'invasions biologiques (MURRAY, 2002; OKUBO ET LEVIN, 2002; SHIGESADA ET KAWASAKI, 1997). Dans ce cadre, le déplacement des individus dans leur environnement est aléatoire et leur trajectoire est modélisée par un mouvement Brownien ou un processus de diffusion stochastique plus général. Les noyaux de dispersion permettent de traduire de façon plus réaliste la redistribution des individus. Ils représentent une fonction de densité de probabilité qu'un individu migre depuis une localisation source (NATHAN *et al.*, 2012). Ils permettent de prendre en compte une plus grande variabilité du processus de dispersion. Ils peuvent notamment inclure, à divers degrés, les mouvements individuels qui mènent à des évènements de dispersion à longue distance (traduit par l'expression Long-Distance Dispersal (LDD) en anglais). On peut distinguer en particulier trois noyaux représentant trois exemples de dispersions caractéristiques (FIGURE 5.2). Le noyau Gaussien modélise une dispersion majoritairement proche de la localisation initiale, avec peu d'évènements de LDD. Le noyau Exponentiel modélise une plus grande fréquence de LDD. Enfin, le noyau Exponentiel-Puissance modélise des évènements de LDD très fréquents. Ce noyau est un noyau dit à queue lourde, *i.e.* pour lequel les évènements de LDD sont plus fréquents que pour un noyau Exponentiel qui partagerait la même distance moyenne de dispersion (KLEIN *et al.*, 2006). Il existe au delà de ces trois exemples un continuum de fonctions de densité pour décrire les cas de dispersion intermédiaire.

La connaissance de la dynamique d'expansion est d'un intérêt primordial pour concevoir des stratégies de gestion fondées sur l'écologie des agents pathogènes et limiter l'impact des maladies émergentes (DYBIEC *et al.*, 2009; FABRE *et al.*, 2021b). Mesurer et inférer la démographie et la dispersion d'une espèce le long d'un front de colonisation peut apporter beaucoup d'informations sur le type de système que l'on étudie et permettre d'anticiper de futures expansions.

5.2 Structures génétiques sur un front de colonisation

Au delà de sa manifestation en terme de dynamique de population, la colonisation peut avoir des conséquences sur l'évolution de la structure génétique au cours de l'expansion. Ces variations de structures génétiques résultent de mécanismes évolutifs. En particulier, quatre de ces mécanismes s'appliquent à l'évolution génétique neutre d'une population (FIGURE 5.3) :

- **Le flux de gènes** représente les variations génétiques dues aux mouvements des individus le long du front de colonisation.
- **La mutation** permet l'apparition aléatoire de variabilité génétique à un endroit donné du front de colonisation.
- **La dérive génétique** peut être particulièrement forte à l'avant du front, où les tailles de populations sont faibles. Un front de colonisation peut ainsi résulter d'effets de fondation successifs par un faible nombre d'individus, diminuant la variabilité génétique de la population à l'avant du front.
- **Le surf génétique** représente l'augmentation en proportion, due au hasard, d'un allèle situé sur le front de colonisation. Ce phénomène est une conséquence directe de la dérive génétique particulièrement forte sur un front de colonisation (EDMONDS *et al.*, 2004; KLOPFSTEIN *et al.*, 2006).

D'autre mécanismes évolutifs peuvent modifier la structure génétique d'une population selon les phénotypes des individus. Un tri spatial des phénotypes de dispersion peut être observé le long du domaine colonisé (MILLER *et al.*, 2020). Ce tri correspond à la distribution spatiale ordonnée des phénotypes de dispersion dans un paysage homogène (SHINE *et al.*, 2011) : les meilleurs disperseurs se retrouvent majoritairement à l'avant du front, tandis que les individus se dispersant moins loin restent à l'arrière du front. De plus, la sélection peut entraîner une évolution génétique au cours de la colonisation, par la reproduction différentielle des individus le long du domaine colonisé selon leur phénotype (et notamment leurs capacités de dispersion).

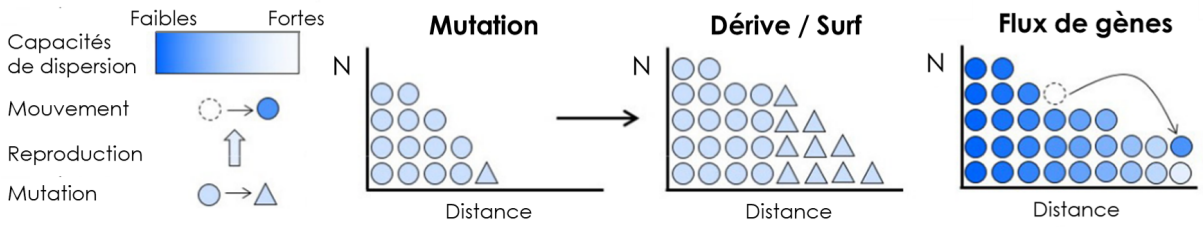


FIGURE 5.3 – Schéma des quatre principaux mécanismes évolutifs impactant la structure génétique neutre d'une population sur un front de colonisation (adapté d'après MILLER *et al.*, 2020).

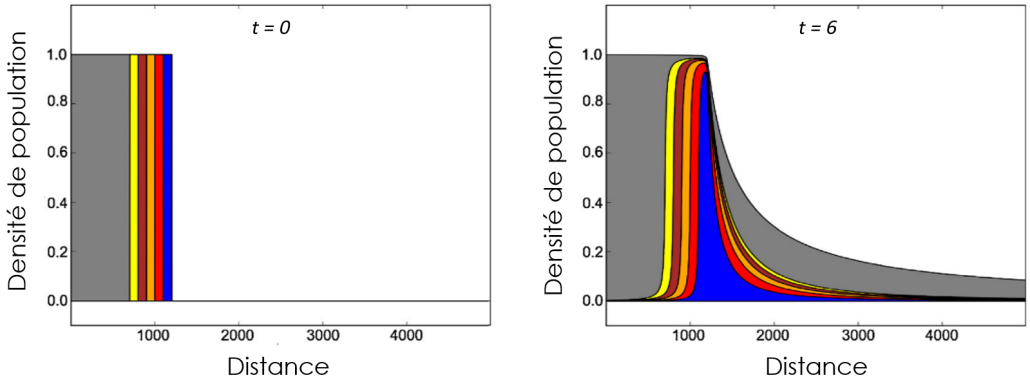
Les mécanismes écologiques gouvernant la dynamique des populations et les mécanismes évolutifs régissant la structure génétique ont été présentés de manière distinctes. Cependant, ces deux mécanismes sont étroitement liés et négliger cette interaction revient à négliger une partie importante du processus. Seule la prise en compte des rétroactions entre les mécanismes écologiques et évolutifs peut permettre de comprendre les structures génétiques qui peuvent émerger le long d'un front de colonisation (KINNISON ET HAIRSTON, 2007; REZNICK *et al.*, 2019).

Une grande partie des travaux de recherche étudiant l'évolution des structures génétiques lors d'une expansion s'intéressent aux processus impliquant la sélection (MILLER *et al.*, 2020). En particulier, de nombreuses études s'intéressent au devenir de mutations favorables ou délétère le long d'un front de colonisation (TRAVIS *et al.*, 2007; BURTON ET TRAVIS, 2008; PEISCHL *et al.*, 2015; GRALKA *et al.*, 2016). L'expansion de l'aire de répartition d'une population peut mener à des structures génétiques très variées, même en l'absence de sélection. Je choisis d'étudier dans cette thèse l'évolution génétique neutre des populations lors d'une colonisation. La compréhension de l'évolution neutre permet de déterminer les attendus évolutifs théoriques et d'interpréter les écarts à cette évolution comme un résultat de la sélection.

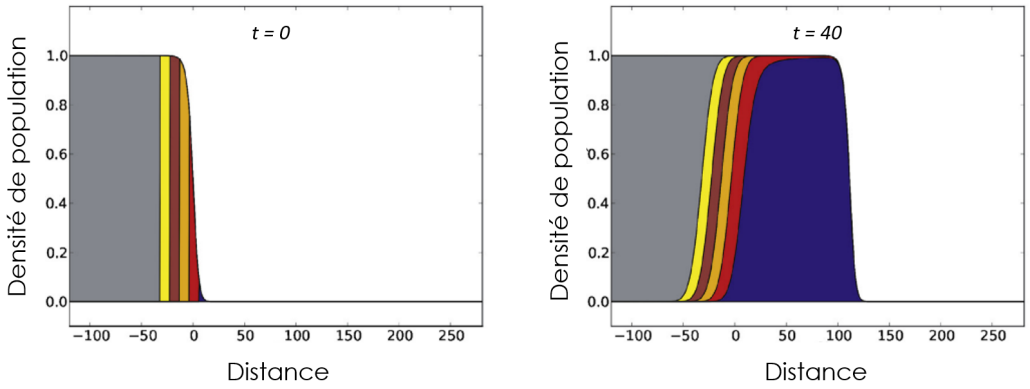
A elles seules, la démographie et la dispersion peuvent fortement impacter les structures génétiques résultant d'une colonisation *via* leurs influences respectives sur la mutation et la dérive génétique et sur les flux de gènes. L'impact de la dispersion sur de telles structures génétiques est double (SLATKIN, 1985; LE CORRE ET KREMER, 1998). D'une part, les flux de gènes réguliers entre les populations *via* les individus migrants favorisent l'homogénéisation génétique des populations. D'autre part, les individus colonisateurs favorisent la différenciation génétique en fondant des sous-populations génétiquement différenciées de la population source (IBRAHIM *et al.*, 1996, AUSTERLITZ *et al.*, 1997, LE CORRE ET KREMER, 1998). Cette dualité peut aboutir à des structures génétiques très différentes sur un front de colonisation, selon la force relative des mécanismes de dispersion en jeu (LE CORRE ET KREMER, 1998).

Deux structures génétiques extrêmes peuvent émerger le long de fronts de colonisation nommés front poussé et front tiré (FIGURE 5.4, BONNEFON *et al.*, 2014). Lors d'un front poussé, tous les individus peuvent contribuer à l'avancée du front de colonisation, quelle que soit leur localisation. Cela maintient une taille efficace de population élevée le long du front (FAYARD *et al.*, 2009). Un front poussé présente donc une conservation stricte de la diversité génétique neutre initiale sur l'ensemble de l'aire géographique colonisée, grâce à un flux de gènes suffisant entre l'arrière du front et l'avant du front (FIGURE 5.4a). A l'inverse, lors d'un front tiré, seuls les individus situés à l'avant du front contribuent à la colonisation. Cela diminue la taille de population efficace et maximise les effets de la dérive génétique à l'avant du front, entraînant des évènements de fondation successifs qui

modifient les fréquences génétiques (EDMONDS *et al.*, 2004; WEI ET KRONE, 2005; HALLATSCHEK *et al.*, 2007). Un front tiré présente donc une érosion progressive de la diversité génétique de l'arrière à l'avant du front (FIGURE 5.4b). Dans les cas extrêmes, ces évènements de fondation peuvent mener à la fixation d'un allèle sur le front de colonisation *via* le processus de surf génétique décrit précédemment (KLOPFSTEIN *et al.*, 2006, HALLATSCHEK *et al.*, 2007,EXCOFFIER ET RAY, 2008).



(a) Front poussé



(b) Front tiré

FIGURE 5.4 – Exemple de l'évolution de la diversité génétique le long d'un front de colonisation pour (a) un front poussé et (b) un front tiré, d'après les simulations théoriques de BONNEFON *et al.*, 2014. Ces fronts résultent des simulations de noyaux de dispersion à (a) forte et (b) faible fréquence d'évènements de LDD. Chaque couleur représente un génotype distinct, t représente le temps passé depuis le début de la colonisation (adapté d'après BONNEFON *et al.*, 2014).

L'un des principaux paramètres de la dispersion distinguant un front poussé d'un front tiré est la fréquence des évènements de LDD (FIGURE 5.5). Pour une forte fréquence de ces évènements, la diversité génétique est conservée le long du front grâce à la migration régulière d'individus de l'arrière vers l'avant du front (ce qui mène à un front poussé, FIGURE 5.4a). Pour une faible fréquence, les évènements de LDD sont insuffisants pour compenser les effets de fondation répétés à l'avant du front de colonisation (ce qui mène

à un front tiré, FIGURE 5.4b). Pour des fréquences intermédiaires, la migration régulière d'individus au delà d'un front de colonisation peut mener à des structures génétiques en taches plus ou moins agrégées le long du front. Dans les cas particuliers où la population colonise un corridor étroit, ces structures en taches peuvent théoriquement mener à un phénomène d'embolie (BIALOZYT *et al.*, 2006). Une embolie se crée lorsque les évènements de LDD sont assez rares et envoient quelques migrants au delà du front de colonisation. Les individus fondateurs se reproduisent alors et forment une population génétiquement peu diverse, voire monomorphe, qui bloque la colonisation des individus provenant du front de colonisation.

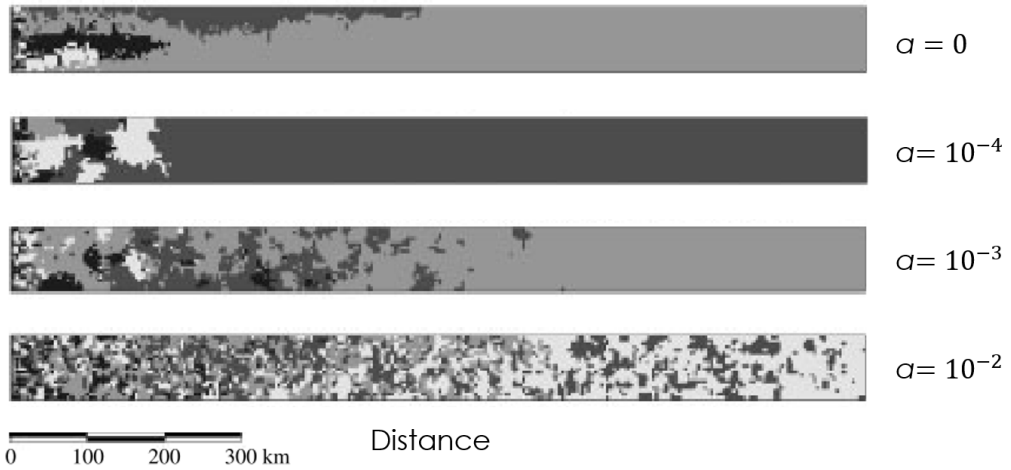


FIGURE 5.5 – Évolution de la diversité génétique le long d'un front de colonisation selon la fréquence des évènements de LDD, d'après les simulations théoriques de BIALOZYT *et al.*, 2006. a représente la fréquence d'évènements de LDD, $a = 0$ correspond à une dispersion sous l'équation de diffusion. Chaque nuance de gris correspond au génotype le plus fréquent dans la cellule modélisée (adapté d'après BIALOZYT *et al.*, 2006).

L'ensemble des processus décrits permet de s'attendre à une grande diversité des structures génétiques neutres que l'on peut observer suite à un évènement de colonisation. Cette diversité peut rendre difficile la caractérisation des évènements de colonisation, notamment dans le cas d'invasions biologiques qui sont des phénomènes uniques. Il est dans ce cas difficile de dissocier ce qui résulte du processus de dispersion et ce qui résulte du hasard qui a pu mener à cette réalisation particulière du processus.

5.3 Cas d'étude : les invasions récurrentes de *Melampsora larici-populina* le long de la vallée de la Durance

La seconde partie de cette thèse se focalise sur un système écologique particulier, la vallée de la Durance, que l'agent pathogène Mlp colonise chaque année d'amont en aval. Cette vallée est encaissée dans les Alpes et la zone la plus en amont, le Briançonnais,

constitue une zone où les aires de répartition géographique du mélèze et du peuplier se chevauchent. Cette zone restreinte est donc favorable à la reproduction sexuée de Mlp sur le mélèze (son hôte écidien), avant le retour par migration sur le peuplier (son hôte télien) au printemps. La vallée de la Durance est bordée par une ripisylve continue de peupliers sauvages (*Populus nigra*) sur plus de 200 km. Mlp colonise chaque année les peupliers bordant cette vallée de l'amont vers l'aval par reproductions clonales successives lors de la phase épidémique (XHAARD *et al.*, 2012). La phase de colonisation se termine en automne lors de la chute des feuilles de peupliers. L'absence de mélèzes en aval ne permet pas aux individus ayant colonisé la vallée de survivre à l'hiver. La vallée de la Durance est donc vidée des agents pathogènes chaque automne, puis re-colonisée depuis la zone de sympatrie entre peupliers et mélèzes chaque année. Cette vallée représente un corridor écologique qui nous permet d'échantillonner des réplicats de colonisations qui ont lieu à partir du même point source, dans une même direction, le long d'un domaine continu en une dimension.

Parmi les trois groupes génétiques de Mlp présents sur le territoire français, seuls les individus appartenant au groupe « Sauvage » contribuent à l'épidémie initiée en amont de la vallée de la Durance (XHAARD *et al.*, 2012). Différents suivis épidémiologiques et génétiques de ces colonisations ont été réalisés dans cette vallée entre les années 2004 et 2011. Une analyse de l'épidémie de l'année 2004 a révélé une structure génétique particulière en fin de colonisation, sans stricte conservation de la diversité génétique neutre de l'amont vers l'aval, mais sans surf génétique pour autant (BECHELER *et al.*, 2016). Au lieu de cela, des lignées clonales ont été échantillonnées à plusieurs endroits de la vallée mais de façon discontinue, comme si les clones avaient "sauté" par dessus des sites durant leur dispersion. Un premier travail de recherche s'est intéressé à décrire cette épidémie (XHAARD, 2011), mais n'a pas permis d'aller jusqu'à l'inférence de la dynamique de dispersion de Mlp. Aucun travail de recherche ne s'est pour l'instant intéressé à la comparaison des différentes épidémies annuelles et des structures génétiques qui en résultent ni au lien entre ces structures et la dynamique d'expansion de l'agent pathogène.

5.4 Objectifs

Dans cette partie, je souhaite mettre en avant le lien entre la dynamique d'expansion et l'évolution génétique d'une population d'agents pathogènes responsables d'une maladie émergente géographique, dans un système pour lequel la composante spatiale est déterminante. J'utilise pour cela l'apport conjoint de la modélisation et des données empiriques sur la colonisation annuelle de Mlp le long de la vallée de la Durance. Cette colonisation est une situation où l'équilibre démographique de la population d'agents pathogènes n'est pas respecté, mais qui représente un système bien décrit permettant de poser de fortes

hypothèses de modélisation : la population source des épidémies est bien identifiée et la colonisation s'opère sur un domaine continu et dans une seule direction. J'utilise ce système particulier pour répondre aux objectifs suivants :

1. Inférer les paramètres de la dynamique d'expansion régissant la colonisation de Mlp dans la vallée de la Durance.
2. Déterminer la variabilité des structures génétiques observées empiriquement au cours de différents évènements de colonisation de Mlp dans la vallée de la Durance.
3. Déterminer l'influence de la dynamique d'expansion de l'agent pathogène sur les structures génétiques résultantes.

Le chapitre 6 répond au premier objectif grâce à un modèle d'inférence par maximum de vraisemblance appliqué aux données de suivi épidémiologique de la colonisation de la valée de la Durance par Mlp en 2008. Le chapitre 7 répond aux objectifs 2 et 3 de cette partie en comparant les structures génétiques spatiales de Mlp en fin d'épidémie pour les années 2004, 2008, 2009 et 2011.

Chapitre 6

Modèle mécanico-statistique et inférence de la dynamique d'expansion de *Melampsora larici-populina* dans la vallée de la Durance

La dynamique d'expansion sur un front de colonisation est en grande partie déterminées par la démographie et le régime de dispersion de la population. Ces deux mécanismes sont rarement considérés conjointement car cela peut s'avérer compliqué d'un point de vue méthodologique. L'objectif de ce chapitre est d'inférer les paramètres régissant ces deux composantes lors de la dynamique d'expansion de Mlp dans la vallée de la Durance.

6.1 Article 5. Inferring invasion determinants with mechanistic models and multitype samples

L'inférence des paramètres régissant la colonisation de Mlp dans la vallée de la Durance repose sur une approche mécanico-statistique qui associe un modèle de processus avec un traitement statistique permettant de retracer les lois d'observation des données. La partie mécaniste modélise une expansion en une dimension selon différents régimes de dispersion. La partie statistique établit les lois d'observation de l'infection en respectant la structure des données obtenues lors du suivi épidémiologique de la colonisation de la vallée de la Durance par Mlp. Enfin, l'étape d'inférence proprement dite couple ce modèle mecanico-statistique aux données empiriques collectées au cours de l'année 2008. Cette approche de modélisation permet d'inférer un ensemble de paramètres de la dynamique d'expansion de Mlp par la méthode du maximum de vraisemblance. Les paramètres se divisent en plusieurs catégories : 1) Les paramètres de la dispersion (la fonction de dispersion, la fréquence des évènements de LDD et la distance moyenne de dispersion) ; 2) les paramètres démographiques (les taux de croissance) ; 3) Les paramètres des lois d'observation de l'infection.

Ces résultats sont ici présentés sous la forme d'un article de recherche en préparation intitulé « Inferring invasion determinants with mechanistic models and multitype samples » et qui sera soumis à la revue METHODS IN ECOLOGY AND EVOLUTION prochainement.

Inferring invasion determinants with mechanistic models and multitype samples

M. Saubin[1], J. Coville[2], C. Xhaard[1,2,3], P. Frey[1], S. Soubeyrand[2], F. Halkett[1*] and F. Fabre[4*]

[1] Université de Lorraine, INRAE, IAM, F-54000 Nancy, France

[2] UR BIOSP, INRAE, Domaine Saint Paul, 84914 Avignon Cedex 9, France

[3] Université de Lorraine, INSERM CIC-P 1433, CHRU de Nancy, INSERM U1116, Nancy, France

[4] INRAE, Bordeaux Sciences Agro, SAVE, F-33882 Villenave d'Ornon, France

[*] These authors co-directed this work

1 Introduction

Dispersal is key in ecology and evolutionary biology (Clobert et al., 2004). From an applied point of view, the knowledge of dispersal is of prime interest for designing ecological-based management strategies in a wide diversity of contexts ranging from the conservation of endangered species to the mitigation of emerging epidemics (Dybiec et al., 2009; Fabre et al., 2021). From a theoretical point of view, the pattern and strength of dispersal sharply impact eco-evolutionary dynamics (*i.e.* the reciprocal interactions between ecological and evolutionary processes). Typically, besides the mean dispersal abilities, the finer features of dispersal have many implications for population dynamics (*e.g.* speed of invasion, metapopulation turnover), genetic structure (*e.g.* gene diversity, population differentiation) and local adaptation (Hallatschek and Fisher, 2014). Mathematically, the movement of dispersers (individuals, spores or propagules for example) can be described by a so-called location dispersal kernel (Nathan et al., 2012) that represents the statistical distribution of the location of the propagules of interest after dispersal from a point source. Since the pioneer works of Mollison (1977), much more attention has been paid to the "fatness" of the dispersal kernel tail. When, at relatively large distance, the shape of the tail of a kernel decreases less or equally slowly than an exponential distribution, the kernel is termed "short-tailed" or "thin-tailed". If it decreases more slowly, the kernel is termed "long-tailed" or "fat-tailed" (Klein et al., 2006) and long-distance dispersal (LDD) events are then more frequent than with an exponential kernel that shares the same mean dispersal distance. Notably, while a species displaying a short-tailed dispersal kernel will generate an invasion front of constant velocity, long-tailed kernels will cause an accelerating front of colonisation (Clark et al., 2001; Kot et al., 1996; Ferrandino, 1993; Mundt et al., 2009; Hallatschek and Fisher, 2014). Frequent LDD events associated with long-tailed kernels can also (i) cause a reshuffling of alleles along the colonisation gradient, which prevent

the erosion of genetic diversity (Nichols and Hewitt, 1994; Petit, 2004; Fayard et al., 2009) or (ii) lead to patchy population structures (Ibrahim et al., 1996; Bialozyt et al., 2006).

Given its biological significance, properly characterising the dispersal kernels of species of interest is a major issue. However, this is a challenging task for many species, especially when dispersing individuals are numerous, small (and thus difficult to track) and move far away (Nathan, 2001; Nathan et al., 2003). This is typically the case of plant pathogens for which very few dispersal kernels are known (Fabre et al., 2021) despite their major impact on food security (Strange and Scott, 2005; Savary et al., 2019). For wind-dispersed propagules such as spores of plant pathogens, but also for pollen or seeds, three approaches can be distinguished to infer dispersal (Nathan et al., 2003; Rieux et al., 2014). The first class of methods uses the indirect effect of dispersal on the pattern of genetic differentiation among populations to estimate dispersal parameters. Most of them only estimate an average dispersal distance (Broquet and Petit, 2009). The second class enables to infer the entire shape of the dispersal kernel by directly observing dispersal with dedicated experiments tracking the movement of tagged propagules (thanks to particular morphological structures or genetic markers). They often rely on heavy experimental procedures for data acquisition requiring, for example, to control source strength and location (Rieux et al., 2014). The third class of methods make use of mathematical modelling to infer dispersal from observations of ongoing invasions (Rieux et al., 2013; Grosdidier et al., 2018). If phenomenological models can hardly disentangle the observed data from the dispersal process *per se*, mechanistic-statistical models do so and enable a proper inference of dispersal using spatio-temporal datasets (Wikle, 2003; Hefley et al., 2017; Roques et al., 2011; Soubeyrand and Roques, 2014). These models allow for the parsimonious representation of both growth and dispersal processes in heterogeneous environments (Papaïx et al., 2022). They require detailed knowledge of the biology of the species of interest to properly model the invasion process. They combine a mechanistic model describing the invasion process and a probabilistic model describing the observation process while enabling a proper inference using spatio-temporal data.

Classically, the dynamics of large populations are well described by deterministic differential equations. Following the seminal work of Fisher on the spread of a mutant gene in a given population (Fisher, 1937), invasions have often been modelled through reaction-diffusion equations (Murray, 2002; Okubo and Levin, 2002; Shigesada and Kawasaki, 1997). In this setting, individuals are assumed to move randomly in their environment. Their trajectories are modelled using a Brownian motion or a more general stochastic diffusion process. Whereas reaction-diffusion equations have been classically used to describe biological invasions, their incorporation into mechanistic-statistical approaches to estimate parameters of interest from spatio-temporal data essentially dates back to the early 2000s. Examples include the study of the invasion processes of the House Finch bird over the eastern United States (Wikle, 2003), the pine processionary moth in France (Roques et al., 2011; Soubeyrand and Roques, 2014) and the wolf in Eastern France (Louvrier et al., 2020). In reaction-diffusion equations, individual dispersal is modelled by a diffusion process rather than explicitly using dispersal kernels as in integro-differential equations. In integro-differential equations, individuals are redistributed ac-

cording to the considered kernel (Hutson et al., 2003; Kolmogorov et al., 1937). Contrary to the reaction-diffusion equation, this class of equations enables to take into account putative LDD events. Moreover, they allow considering a large variety of dispersal functions, typically with either short or long tail, rendering this approach very flexible for inference purposes and more likely to model accurately the true organism's dispersal process. However, as far as we know, integro-differential equations have rarely been embedded into mechanistic-statistical approaches to infer dispersal processes in ecology (but see Szymańska et al., 2021 for a recently proposed application of non-local model to cell proliferation).

Data acquisition is another challenge faced by biologists in the field, all the more that data confined to relatively small spatial scales can blur the precise estimates of the shape of the kernel's tail (Ferrandino, 1996). Indeed, the spatial scale at which observations are realised is a major concern when fitting and comparing dispersal kernels. Kuparinen et al. (2007) and Rieux et al. (2014) showed how kernels with very different tails may yield similar fits. They found that the predicted dispersal at long distances depends on both the kernel considered and the distances over which the dispersal data was collected. Accordingly, to gather as much information as possible, it is mandatory to collect data over a wide range of putative population sizes (from absence to near saturation) along the region of interest. This change of scale then implies inferring not only the dispersal kernel but other parameters such as growth rates. This leads to epidemiological models that go beyond simply fitting dispersal gradients to biological data. In practice, such a sampling scheme may be difficult to grasp from a single data type as one faces a trade-off in data acquisition between a large-scale monitoring (*e.g.* the screening of a large number of putative habitat units to test for presence/absence) and, when present, a small-scale observation (*e.g.* the focusing on a reduced number of habitat units to estimate more precisely local population sizes and individual aggregation). Both data types may bring complementary information and could be combined to increase inference accuracy of spatial ecological processes (Gotway and Young, 2002). The probabilistic model describing the observation process in the mechanistic-statistical approach can handle such multiple datasets with varying supports (Wikle, 2003). However, inference based on multi-type data remains a challenging statistical issue as the observation variables describing each data type follow different distribution laws (Chagneau et al., 2011) and can be correlated or, more generally, dependent (Georgescu et al., 2014). This imposes to minutely define the conditional links between the observed variables and the model parameters (the so-called observation laws) to identify and examine the complementarity and the putative redundancy between data types.

In this article, we aim to provide a sound inferential framework to estimate dispersal from ecological invasion data using both reaction-diffusion and integro-differential equations. Both formalisms were embedded within the same inferential framework. For the sake of simplicity, our model applies to one-dimension invasion dynamics. In the following, we define the two classes of mechanistic invasion models, establish the observation laws corresponding to raw and refined samplings, and propose a maximum-likelihood method for estimating the model parameters. To confirm that each model parameter can indeed be efficiently estimated given the amount of data at hand (see Soubeyrand and

Roques, 2014), we perform a simulation study to check model parameters' identifiability given a proposed experimental design. We also aim to assess the confidence level in the choice of the dispersal function as derived by model selection. Last, this inferential framework is applied to original ecological data describing the annual invasion of a tree pathogen (*Melampsora larici-populina*, a fungus species responsible for the poplar rust disease) along the riparian stands of wild poplars bordering the Durance River valley in the French Alps (Xhaard et al., 2012).

2 Modelling one-dimensional invasion and observation processes

2.1 A class of deterministic and mechanistic invasion models

In this study, we model the dynamics of a population density $u(t,x)$ at any time t and point x during an invasion using two types of spatially heterogeneous deterministic models allowing to represent a wide range of dispersal processes. Specifically, we considered a reaction-diffusion model (R.D.) and an integro-differential model (I.D.):

$$\text{R.D.}\begin{cases}\partial_t u(t,x) = D\partial_{xx}u(t,x) + r(x)u(t,x).\left(1-\frac{u(t,x)}{K}\right),\\ u(0,x)=u_0(x),\end{cases}\qquad \text{I.D.}\begin{cases}\partial_t u(t,x) = \int_{-R}^{R} J(x-y)[u(t,y)-u(t,x)]\,dy + r(x)u(t,x).\left(1-\frac{u(t,x)}{K}\right),\\ u(0,x)=u_0(x).\end{cases}$$

where t varies in $[0,T]$ (*i.e.* the study period) and x varies in $[-R,R]$ (*i.e.* the study domain). Both equations exhibit the same structure composed of a diffusion/dispersal component and a reaction component. The reaction component, $r(x)u(t,x)\left(1-\frac{u(t,x)}{K}\right)$ in both equations, is parametrised by a spatial growth rate $r(x)$ that takes into account macro-scale variations of the factors regulating the population density and K the carrying capacity of the environment. It models population growth. The diffusion/dispersal component models population movements either by a diffusion process ($D\partial_{xx}u$ in R.D.) parametrised by diffusion coefficient D or by a dispersal kernel (J in I.D.) depending on a specific set of parameters (see below). Let us emphasise that the integro-differential models provide a rich class of possible movement processes. They are particularly relevant in ecology because they can be designed, for example, to model long-distance dispersal. To cover a large spectrum of possible dispersal processes, we use the following parametric form for the kernel J:

$$J := \frac{\tau}{2\alpha\Gamma\left(\frac{1}{\tau}\right)} e^{-\left|\frac{z}{\alpha}\right|^{\tau}}$$

with mean dispersal distance $\lambda := \alpha\frac{\Gamma\left(\frac{2}{\tau}\right)}{\Gamma\left(\frac{1}{\tau}\right)}$. Varying the value of τ leads to kernels classically used in dispersal studies. Specifically, J can be a Gaussian kernel ($\tau=2, \lambda=\alpha/\sqrt{\pi}$), an exponential kernel ($\tau=1,\lambda=\alpha$) or a fat-tail kernel ($\tau<1, \lambda=\alpha\Gamma\left(\frac{2}{\tau}\right)/\Gamma\left(\frac{1}{\tau}\right)$). Explicit formulas for the solution $u(t,x)$ of these reaction-diffusion/dispersal equations being out of reach, we compute a numerical approximation u_{num} of u which serves as a surrogate for the real solution. Details of the numerical scheme used to compute u_{num} can be found in Appendix A.

2.2 A conditional stochastic model to handle micro-scale fluctuations

Among the factors driving population dynamics, some are structured at large spatial scales (macro-scale) and others at local scales (micro-scale). Both scales must be considered when studying biological invasions. In the model just introduced, the term $r(x)$ describes factors impacting population growth rate at the macro-scale along the whole spatial domain considered. Accordingly, the function $u(t,x)$ is a mean-field approximation of the true population density at macro-scale. Furthermore, the population density can fluctuate due to micro-scale variations of other factors regulating population densities (*e.g.* because of variations in the micro-climate and the host susceptibility). To take into account such local fluctuations, the deterministic invasion model is combined with a conditional probability distribution of the local population size in a habitat unit. The habitat unit is typically a small fraction of the total habitat. This distribution is conditional on $u(t,x)$, the macro-scale population density, and depends on the (unobserved) suitability of the habitat unit. The following paragraphs detail this dependence.

Consider a habitat unit i whose centroid is located at x_i, and suppose that the habitat unit is small enough to reasonably assume that $u(t,x) = u(t,x_i)$ for all location x in the habitat unit. Let $N_i(t)$ denote the number of individuals in i at time t. The conditional distribution of $N_i(t)$ is modelled by a Poisson distribution:

$$N_i(t) \mid u(t,x_i), R_i(t) \sim Poisson(u(t,x_i)R_i(t)), \tag{1}$$

where $R_i(t)$ is the intrinsic propensity of the habitat unit i to be occupied by individuals of the population at time t. Thereafter, $R_i(t)$ is called habitat suitability and takes into account factors like the exposure and the favorability of habitat unit i. The suitability of habitat unit i is a random effect (unobserved variable) and is assumed to follow a Gamma distribution with shape parameter σ^{-2} and scale parameter σ^2:

$$R_i(t) \sim Gamma(\sigma^{-2}, \sigma^2). \tag{2}$$

This parameterisation implies that the mean and variance of $R_i(t)$ are 1 and σ^2, respectively; that the conditional mean and variance of $N_i(t)$ given $u(t,x_i)$ are $u(t,x_i)$ and $u(t,x_i) + u(t,x_i)^2\sigma^2$, respectively; and that its conditional distribution is:

$$N_i(t) \mid u(t,x_i) \sim Negative - Binomial\left(\sigma^{-2}, \frac{u(t,x_i)\sigma^2}{1+u(t,x_i)\sigma^2}\right). \tag{3}$$

2.3 Multi-type sampling and models for the observation processes

During an invasion, the population density may range from zero (beyond the front) to the maximum carrying capacity of the habitat. Thus, it may be relevant to carry out different sampling procedures depending on the population density at the sampling sites. In this article, we consider a two-stage sampling made of one raw sampling which is systematic and one optional refined sampling. In our biological application, we consider the invasion of a foliar pathogenic fungus infecting poplar trees and responsible for rust disease. Therefore, a habitat unit is a poplar leaf. For the raw sampling, we prospected each site at each date to search for rust disease by inspecting leaves in randomly distributed poplar trees. Each leave inspected on the same tree during the raw sampling comes from a

different twig. The refined sampling consisted in randomly sampling twigs on different trees and recording, for each, the total number of leaves and the number of infected leaves. This vocabulary (leaves, twigs and trees) is adapted to our case study for the sake of clarity, but a wide variety of multi-type sampling strategies can be defined and implemented in the model, as long as it follows a two-stage sampling strategy. The following provides details about the multi-type sampling adapted to the case study tackled in this article.

The fungal population is monitored in sampling sites $s \in \{1, \ldots, S\}$ and at sampling times $t \in \{t_1, \ldots, t_K\}$. Sampling sites are assumed to be small with respect to the study region and the duration for collecting one sample is assumed to be short with respect to the study period. Thus, the density of the population at sampling time t in sampling site s is constant and equal to $u(t, z_s)$ where z_s is the centroid of the sampling site s.

Any sampling site s is assumed to contain a large number of leaves which are, as a consequence of the assumptions made above, all associated with the same population density function: $u(t, x_i) = u(t, z_s)$ for all leaves i within sampling site s. To obtain a representative assessment of the population density in any sampling site s, several observations are made in s at each sampling date. Each observation is made in a so-called sampling unit. As we will see below, a sampling unit is either a group of independent leaves (tree) or a group of connected leaves (twig).

In our case study, nearby leaves often encounter the same environmental conditions and, therefore, are characterised by similar habitat suitabilities. This spatial dependence was taken into account by assuming that the leaves of the same twig (considered as a small group of spatially connected leaves) share the same leaf suitability. Accordingly, suitabilities are considered as shared random effects.

In the first stage of the sampling procedure (also called raw sampling), in each sampling site s and at each sampling time t, a number B_{st} of trees is monitored for the presence of infection. We then count the number of infected trees Y_{st} among the total number B_{st} of observed trees. In the simulations and the case study tackled below, the random variables Y_{st} given $u(t, x_s)$ are independent and distributed under the conditional probability distribution f_{st}^{raw} describe in Appendix B.2. The distribution f_{st}^{raw} is a Binomial distribution. It integrates in its success probability a parameter γ describing how the probabilities of leaf infection perceived by the person in charge of the sampling differ between trees from true probabilities (as informed by the mechanistic model) due, typically, to the specific configuration of each tree.

Optionally, depending on disease prevalence and available time, the second stage of the sampling procedure (also called refined sampling) is carried out in site s at time t. In this case, in site s and at time t, G_{st} twigs are collected. For each twig g, we count the total number of leaves M_{stg} and the number of infected leaves Y_{stg}. In the simulations and the case study tackled below, the random variables Y_{stg} given $u(t, x_s)$ are independent and distributed under conditional probability distributions denoted by f_{st}^{ref} described in Appendix B.3. The distribution f_{st}^{ref} is a new mixture distribution (called Gamma-Binomial distribution) obtained using Equations (1–2) and taking into account the spatial dependence of unobserved suitabilities (see Appendix B.3).

2.4 Coupling the mechanistic and observation models

The submodels of the population dynamics and the observation processes described above can be coupled to obtain a mechanistic-statistical model (also called physical-statistical model; Berliner, 2003; Soubeyrand et al., 2009) representing the data and depending on dynamical parameters, namely the growth and dispersal parameters. The likelihood of this mechanistic-statistical model can be written:

$$L(\theta) = \prod_{s=1}^{S} \prod_{t=t_1}^{t_K} \left\{ f_{st}^{raw}(Y_{st}) \left(\prod_{g=1}^{G_{st}} f_{st}^{ref}(Y_{stg}) \right)^{\mathbf{1}(\mathbf{Y_{st}} > \bar{\mathbf{y}})} \right\}, \tag{4}$$

where expressions of f_{st}^{raw} and f_{st}^{ref} adapted to the case study tackled below are given by Equations (S10) and (S14) in Appendix B. The power $\mathbf{1}(\mathbf{Y_{st}} > \bar{\mathbf{y}})$ (that is equal to 1 if $Y_{st} > \bar{y}$, 0 otherwise) implies that the product $\prod_{g=1}^{G_{st}} f_{st}^{ref}(Y_{stg})$ only appears if the refined sampling is carried out in site s. Besides, such a product expression for the likelihood is achieved by assuming that leaves in the raw sampling and those in the refined sampling are not sampled from the same trees. If this does not hold, then an asymptotic assumption like the one in Appendix B.2 can be made to obtain Equation (4), or the dependence of the unobserved suitabilities must be taken into account and another likelihood expression must be derived.

3 Parameter estimation and model selection

We performed simulations to check the practical identifiability of several scenarios of biological invasions. Invasion scenarios represent a wide range of possible states of nature regarding the dispersal process, the environmental heterogeneity at macro-scale and the intensity of local fluctuations at micro-scale. Even though the simulations are designed to cope with the structure of our real data set (Appendix D), the results enable some generic insights to be gained. Specifically, we considered six sampling dates evenly distributed in time, and 12 samplings sites evenly distributed in space. For each pair *(date, site)*, we simulated the raw sampling of 100 trees and the refined sampling of 20 twigs. For the fifth sampling date, the raw sampling was densified with 45 sampling sites instead of 12.

The simulations firstly explored four hypotheses J_{Exp}, J_{Gauss}, J_{ExpP} and RD for the dispersal process. Hypotheses J_{Exp} and J_{Gauss} state that individuals dispersed according to Exponential and Gaussian kernels, respectively, with parameter $\theta_J = (\lambda)$. Hypothesis J_{ExpP} states that individuals dispersed according to a fat-tail Exponential-power kernel with parameters $\theta_J = (\lambda, \tau)$ and $\tau < 1$. Finally, hypothesis RD states that individual dispersal is a diffusion process parameterised by $\theta_J = (\lambda)$. The parameter λ represents the mean distance travelled whatever the dispersal hypothesis considered. Simulations secondly considered environmental heterogeneity at macro-scale by varying the intrinsic growth rate of the pathogen population (r) in space. Specifically, along the one-dimensional domain, we considered 2 values of r parameterised by $\theta_r = (r_{dw}, \omega)$ such that $r_{up} = r_{dw}e^{\omega}$. Finally, simulations take into account the environmental heterogeneity at the micro-scale by varying the parameter of leaf suitability σ^2, and tree perception γ. Thereafter, $\theta = (\theta_r, \theta_J, \gamma, \sigma^2)$ denotes the vector of model parameters.

3.1 Accurate inference of model parameters

To check if the data sets collected are informative enough to efficiently estimate the parameters of a given model (the so-called practical identifiability), we proceeded in 3 steps for each dispersal hypothesis : (i) a set of parameter values $\theta = (\theta_r, \theta_J, \gamma, \sigma^2)$ is drawn from dedicated distributions that encompass a large diversity of realistic invasions, (ii) a data set with a structure similar to our real sampling is simulated given θ and (iii) θ is re-estimated using maximum-likelihood method applied to the data set. These steps were iterated until the acceptance of $n = 100$ data sets describing realistic invasions. Details on the simulation procedure and conditions for realistic invasions are provided in Appendix D.1.

Practical identifiability was tested by means of correlation coefficients between the true and estimated parameter values for four models corresponding to the four dispersal processes (Table 1, Figures S1, S2, S3, S4). Correlation coefficients and standard deviations are given for natural scale for parameter ω, and for logarithm scales for parameters r_{dw}, γ, λ, τ, and σ^2.

All the parameters defining the mechanistic invasion model (r_{dw}, ω, λ) displayed very good practical identifiability with correlation coefficients above 0.93. In the case of the Exponential-power dispersal kernel, the additional parameter representing the tail of the distribution (τ) also displayed a very good practical identifiability with a correlation coefficient of 0.95. Regarding the parameters defining the observation model, the local fluctuations, σ^2 displayed particularly high correlation coefficients (0.99 for all models), while γ also displayed good, but somewhat lower, practical identifiability for all models (above 0.83).

Table 1: Model practical identifiability. Numbers indicate the coefficient of correlation between the true and estimated parameter values for four models corresponding to the four dispersal processes (J_{Exp}, J_{Gauss}, J_{ExpP} and RD) on 100 replicates. High correlation between true and re-estimated parameters indicates a good practical identifiability. The standard deviations of the coefficients of correlation, estimated with a bootstrapping method, are indicated in brackets. Correlation coefficients and standard deviations are given for natural scale for parameter ω, and logarithm scales for parameters r_{dw}, γ, λ, τ, and σ^2

	J_{Exp}	J_{Gauss}	J_{ExpP}	RD
r_{dw}	$0.99(1.10^{-3})$	$0.99(1.10^{-3})$	$0.99(2.10^{-3})$	$0.93(6.10^{-2})$
ω	$0.99(< 10^{-3})$	$0.99(< 10^{-3})$	$0.99(1.10^{-3})$	$0.99(1.10^{-3})$
λ	$0.99(5.10^{-3})$	$0.98(8.10^{-3})$	$0.99(1.10^{-3})$	$0.95(2.10^{-2})$
τ			$0.95(1.10^{-2})$	
γ	$0.85(4.10^{-2})$	$0.83(4.10^{-2})$	$0.83(5.10^{-2})$	$0.84(3.10^{-2})$
σ^2	$0.99(1.10^{-3})$	$0.99(< 10^{-3})$	$0.99(< 10^{-3})$	$0.99(< 10^{-3})$

3.2 Confidence in the selection of the dispersal process

Numerical simulations were next designed to test whether model selection can disentangle the true dispersal process (*i.e.* the dispersal hypothesis used to simulate the data set) from alternative dispersal processes. We proceeded as indicated for model practical identifiability, except that we fitted to each data set the true model (as previously) but also the 3 other models corresponding to alternative dispersal hypotheses until acceptance of $n = 50$ replicates. Models were ranked using AIC (Akaike

Information Criteria, see Appendix D.2) to select the model best supported by the data set. Confidence level in model selection was assessed by the proportion of cases where the true model was selected according to AIC (Table 2). Recall that, as we only consider $\tau < 1$ in our numerical simulations (Appendix D.1), the dispersal kernel J_{ExpP} is always a fat-tail kernel.

The model selection procedure is efficient for the dispersal hypotheses Exponential-power J_{ExpP}, Exponential J_{Exp}, and reaction-diffusion RD, with 70%, 62% and 58% of correct kernel selection, respectively (Table 2). When the fat-tail Exponential-power kernel is not correctly identified, it is mostly mistaken with the Exponential one (for 20% of the simulations). In line with this, the probability of correctly selecting the kernel J_{ExpP} decreases when the parameter τ increases towards 1, the value for which the Exponential-power kernel corresponds to the Exponential kernel (Figure 1). Importantly, when the Exponential-power kernel is correctly selected, we observe a large difference between its AIC and the AIC of the second best model (89.62 points on average). Conversely, when the invasion process is simulated under J_{ExpP}, but another kernel is selected, we observe a very small AIC difference (0.38 point on average). Model selection does not allow to correctly select the Gaussian kernel J_{Gauss} (Table 2). Indeed, with only 26% of correct model selection, this kernel is not better identified than with a random draw of one of the four models, which would lead to 25% of correct estimations. Its correct identification is greatly improved by densifying the sampling scheme (Appendix D.5, Table S3). Finally, note that, when the invasion process is simulated under model RD or J_{Gauss}, a short-tail kernel is always selected and, thus, never confounded with the fat-tail kernel J_{ExpP}.

Table 2: Efficiency of model selection using Akaike information criterion (AIC). The four first columns indicate the proportion of cases, among 50 replicates, where each tested model was selected using AIC, given that data sets were generated under a particular model (*i.e.* true model). Column $dAIC_{true}$ (*resp.* $dAIC_{wrong}$) indicates the mean difference between the AIC of the model selected when the model selected is the true one (*resp.* when the model selected is not the true model) and the second best model (*resp.* being the true model or not).

	Selected Model					
True Model	J_{Exp}	J_{Gauss}	J_{ExpP}	RD	$dAIC_{true}$	$dAIC_{wrong}$
J_{Exp}	**0.62**	0.22	0.06	0.10	0.84	0.74
J_{Gauss}	0.34	**0.26**	0.00	0.40	1.08	0.55
J_{ExpP}	0.20	0.04	**0.70**	0.06	**89.62**	0.38
RD	0.18	0.24	0.00	**0.58**	0.71	0.23

4 Case study: Invasion of poplar rust along the Durance River valley

4.1 Study site

We applied our approach to infer the dispersal of the plant pathogen fungus *Melampsora larici-populina*, responsible for the poplar rust disease, from the monitoring of an epidemic invading the Durance River valley. Embanked in the French Alps, the Durance River valley constitutes a one-dimension ecological

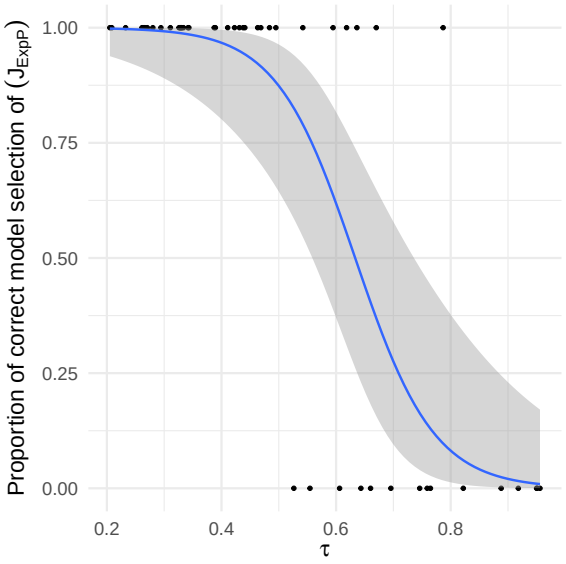


Figure 1: Logistic regression on the proportion of correct model selection of dispersal J_{ExpP} as a function of τ. Dots represent the values of τ used for the 50 replicates of simulated dispersal model J_{ExpP}, and the estimated dispersal model (1 for a correct model selection of J_{ExpP} and 0 for an wrong model selection). The blue line corresponds to the predicted value of the proportion of correct model selection J_{ExpP} as a function of τ, and the grey area corresponds to the confidence band at 95%.

corridor which channels annual epidemics of the poplar rust pathogen *M. larici-populina*. Each year the fungus has to reproduce on larches (*Larix decidua*) that are located in the upstream part of the valley only. This constitutes the starting point of the annual epidemics. Then the fungus migrates to poplar leaves and performs several rounds of infection until leaf-fall. Each infected leaf produces thousands of spores that are wind-dispersed. All along the valley, the Durance River is bordered by a nearly continuous riparian forest of wild poplars (*Populus nigra*). The annual epidemic on poplars thus spreads downstream through the riparian stands, mimicking a one-dimension biological invasion (Xhaard et al., 2012). A previous genetic study showed that the epidemic was indeed initiated in an upstream location where poplars and larchs coexist (Prelles) and progresses along the valley (Becheler et al., 2016). In fall, the corridor is cleared for disease after leaf-fall. At 62 km downstream of the starting point of the epidemics, the Serre-Ponçon dam represents a shift point in the valley topology, with a steed-sided valley upstream and a larger riparian zone downstream. This delimitation led us to consider 2 values of growth rates r along the one-dimensional domain: r_{up} and r_{dw} (see Appendix D for details).

4.2 Monitoring of an annual epidemic wave

In 2008, rust incidence was monitored every three weeks from July to November at 12 sites evenly distributed along the valley (Figure 2). Sites were inspected during seven rounds of surveys. For a unique date (Oct. 22), the raw sampling was densified with 45 sites monitored instead of 12. We focused on young poplar trees (up to 2m high) growing on the stands by the riverside. Two monitorings

were conducted, the raw and refined sampling, respectively. For the raw sampling, we prospected each site at each date to search for rust disease by inspecting randomly distributed poplar trees (different trees at different dates for a given site). Depending on rust incidence and poplar tree accessibility, 40 to 150 trees (mean 74) were checked for disease. Each tree was inspected through a global scan of the leaves on different twigs until at least one infected leaf was found or after 30 s of inspection. The tree was denoted infected or healthy, respectively. This survey method amounts to minutely inspecting 10 leaves per tree, *i.e.* with the same efficiency of disease detection as through the refined sampling (see details of the statistical procedure in Appendix C). The global scan procedure of the trees leads to equivalently surveying fewer and fewer leaves as the epidemic progresses. Optionally, when at least one tree was infected, and depending on available time, we carried out a refined sampling to collect more information on the variance in disease susceptibility (*i.e.* habitat suitability) among the sampling domain. The refined sampling consisted in randomly sampling 20 twigs on different trees and recording, for each, the total number of leaves and the number of infected leaves.

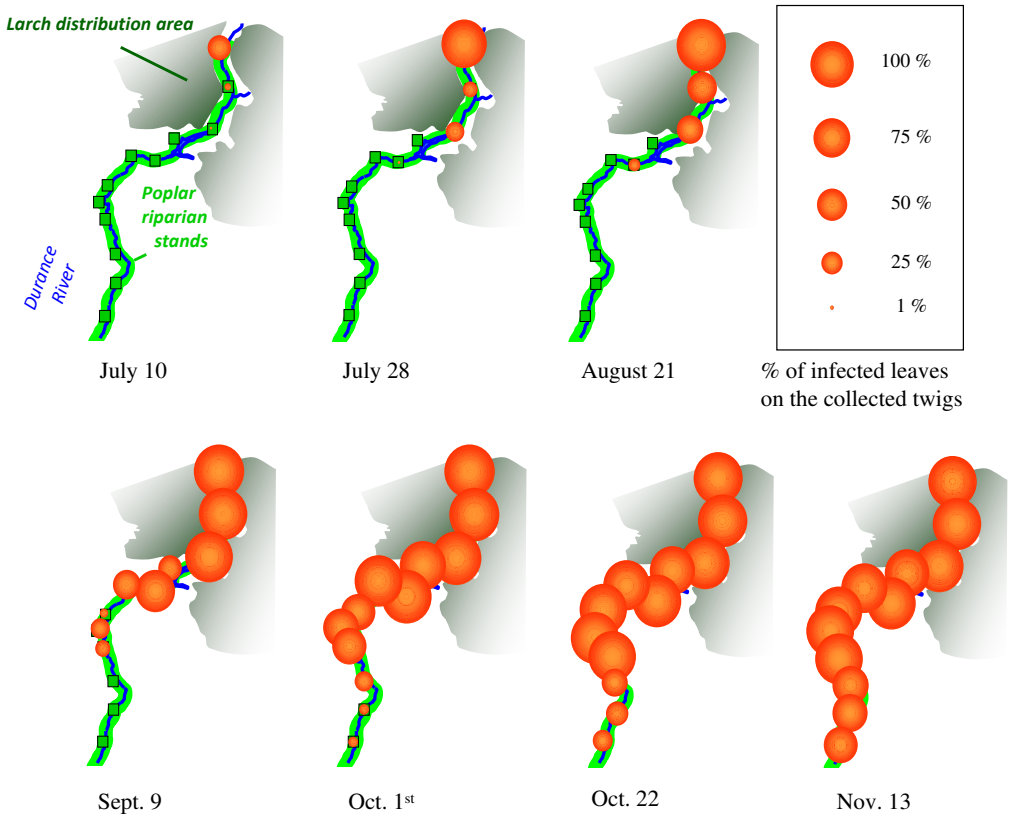


Figure 2: Poplar rust epidemic wave along the Durance River valley in 2008. The larch distribution area is represented in dark green, wild poplar riparian stands in pale green. The 12 study sites are represented by the green squares. Orange dots describe the evolution of the poplar rust epidemic through time (7 rounds of disease notation) and space (12 studied sites). Dot size is proportional to rust disease incidence assessed from the refined sampling.

4.3 Dispersal and demographic processes ruling the epidemic wave

In our case study, $u(t, x_s)$ is the density of fungi infection at time t at point x on a poplar leaf. Each leaf represents a habitat unit (carrying a maximum of 750 lesions, Appendix E). In the raw sampling, trees are inspected. A tree represents a habitat block. The probability p_{st} that a tree is infected is given by the probability that at least one leaf is infected among the set of M leaves observed (Eq. (S9), Appendix B.2). In the refined sampling, twigs are inspected. A twig is a group of spatially connected leaves sharing the same susceptibility to rust disease (*i.e.* the same habitat suitability). Different twigs have different habitat suitability $\mathcal{R}_g(t)$ because of local environmental variations resulting, for example, from their position and/or orientation in the tree.

Model selection was used to decipher which dispersal process was best supported by the data set, for five initial conditions. The large AIC difference in favour of hypothesis J_{ExpP} indicates that poplar rust propagules assuredly disperse according to an exponential-power dispersal kernel along the Durance River valley (Table 3). Note that for all kernels, the five initial conditions lead to similar estimations. Under the RD hypothesis, however, initial conditions can lead to different estimations because of local optima, but all AIC resulting from the RD hypothesis are higher than AIC resulting from the three dispersal kernels.

Table 3: Model selection for the epidemic of poplar rust along the Durance River valley. The Akaike information criteria are indicated for each model fitted to the real data set. The model best supported by the data is indicated in bold. AIC_{mean}, AIC_{median}, AIC_{min}, and AIC_{max} represent summary statistics among the AIC of all replicates: mean, median, minimum, and maximum, respectively.

Dispersal	AIC_{mean}	AIC_{median}	AIC_{min}	AIC_{max}
J_{Exp}	5476	5476	5476	5477
J_{Gauss}	5509	5510	5508	5510
$\boldsymbol{J_{ExpP}}$	**5180**	**5179**	**5179**	**5182**
RD	6596	6303	6303	7769

Parameter's estimation for the best model are summarised in Table 4. The parameters of the Exponential-power kernel are of prime interest in our study. They firstly indicate that the mean distance travelled by rust spores is estimated at 2.02 km. Moreover, the mean exponent parameter τ of the Exponential-power kernel is 0.24. This value, much lower than 1, suggests substantial long-distance dispersal events. We also estimated the growth rates of the poplar rust epidemics along the Durance River valley. From upstream to downstream, their mean estimates are 0.086 and 0.020, respectively. Lastly, the mean two parameters of the observation model, γ and σ^2, amount to 5.07 and 1.1, respectively. The parameter γ represents how perceived probabilities of leaf infection differ among trees from true probabilities. If its estimated value 5.07 indicates some variability in the perception of infected leaves, this variability is moderate as the shape of the underlying Beta-Binomial distribution approaches the Binomial distribution (for which perception differences are absent) (figure 3, line 1). By contrast, the estimated value of $\sigma^2 = 1.1$ suggests a substantial variability in leaf suitability between twigs. This is evidenced by comparing the shape of the estimated Gamma-Binomial distribution with a situation with negligible differences in receptivity between twigs (Figure 3, line 2, case $\sigma^2 = 0.01$).

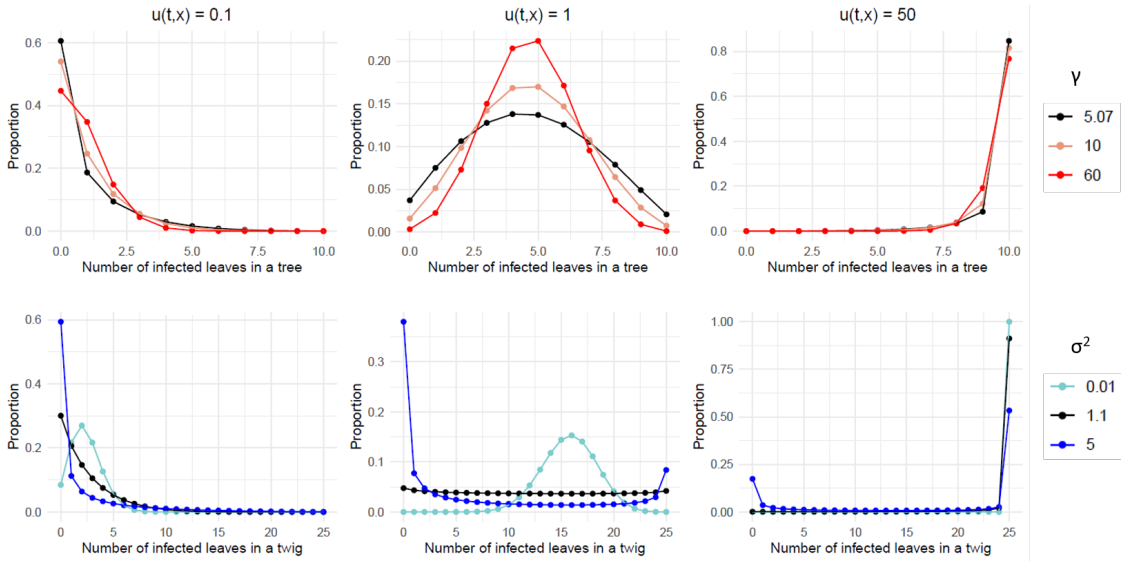


Figure 3: Distributions of the number of infected leaves in a tree and of the number of infected leaves in a twig, for increasing densities of infection $u(t,x)$, and contrasted levels of environmental heterogeneity σ^2 and γ. The number of infected leaves in a tree follows a Beta-Binomial distribution (Eq. (S8)) with $\sigma^2 = 1.1$. Its density is plotted for three tree perceptions γ: 5.07 (estimated value on the real data set), 10 (intermediate value) and 60 for which the Beta-Binomial distribution is approaching a Binomial distribution. The number of infected leaves in a twig follows a Gamma-Binomial distribution (Eq. (S14)). Its density is plotted for three leaf suitabilities σ^2: 1.1 (estimated value on the real data set), 5 (a higher value) and 0.01 a value lowering variability in leaf suitability between twigs (when σ^2 tends to 0, all twigs share the same leaf suitability).

Table 4: Statistical summary of the inference of the parameters for the model best supported by the real data set J_{ExpP}. Summary statistics are calculated from the method detailed in Appendix D.3.

Parameter	*Description*	$q-2.5\%$	*median*	*mean*	$q-97.5\%$
r_{up}	Growth rate upstream	0.0312	0.0857	0.0859	0.191
r_{dw}	Growth rate downstream	0.0114	0.0202	0.0202	0.0289
λ	Mean dispersal distance	1.76	2.01	2.02	2.27
τ	Kernel exponent parameter	0.220	0.240	0.240	0.263
γ	Tree perception	3.21	5.04	5.07	6.77
σ^2	Leaf suitability	0.987	1.10	1.10	1.21

Model check consists in testing whether the selected model was indeed able - given the parameter values inferred above - to reproduce the observed data describing the epidemic wave that invaded the Durance River valley in 2008. To do so, we assessed the coverage rate of the raw sampling data (proportions of infected trees) based on their 95%-confidence intervals (Appendix D.4, Figure 4). Over all sampling dates, the mean coverage rate is high (0.74), which indicates that the model indeed captures a large part of the strong variability of the data.

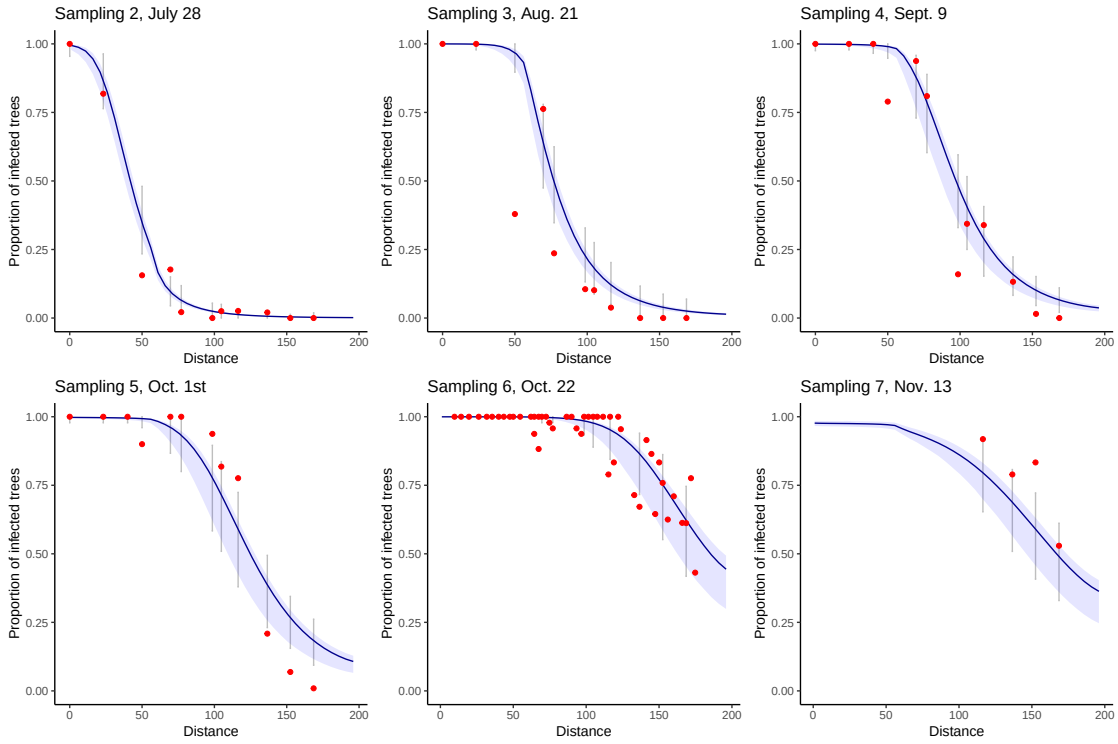


Figure 4: Model check: Coverage rates for the raw sampling. Each sampling date is represented on a separate graph. Sampling 1 is not represented because it corresponds to the initial condition of the epidemics for all simulations. Blue areas correspond to the 95% confidence intervals on the proportion of infected trees, grey intervals correspond to the 95% prediction intervals at each site, *i.e.* taking into account the observation laws given the proportion of infected trees. Red points correspond to the observed data. Only four observations are available for sampling 7 because at this date (November 13) the leaves had already fallen from the trees upstream.

5 Discussion

This study originally combines mechanistic and statistical modelling applied to count data, with the aim to infer demographic parameters during a biological invasion. A strength of the mechanistic model was to consider population growth combined with different dispersal regimes with the aim to grasp all types of processes that may be involved in colonisation. The modelling approach was enriched with a sound statistical treatment that considers different types of count data which broaden our approach and increases the confidence in our estimates. Although it is generic, our model was tuned to fit a precise survey: the annual spread of the poplar rust fungus *M. larici-populina* along the Durance River valley. This valley channels every year the spread of an epidemic along a one-dimensional corridor of nearly 200 km (Xhaard et al., 2012; Becheler et al., 2016). The monitoring we performed enables to build a comprehensive data set at a large spatial scale, which is mandatory to precisely infer the shape of the tail of dispersal kernels (Ferrandino, 1996; Kuparinen et al., 2007). To this end, we consider several demographic scenarios with a particular focus on the characterisation of the dispersal processes. We demonstrate how reliable our approach is in assessing the confidence in the model choice and estimating parameters. A main originality in the statistical treatment of the data was to consider that

habitat suitability can vary over the sampling domain. This encapsulates a large part of the variability in the biological processes and provides robust estimates of the parameters of biological interest.

5.1 Estimation of the dispersal kernel of the poplar rust

This study provides the first reliable estimation of dispersal in the fungal plant pathogen *M. larici-populina*. Dispersal kernels are firstly defined by their scale, which can be taken to correspond to the mean dispersal distance. The mean transmission distance obtained from the best model is 2.02 km with a 95% confidence interval ranging from 1.76 to 2.27 km. Although few such estimates are available, the mean dispersal distance is likely much higher for fungal plant pathogens than for insect-transmitted pathogens. For example, a (non-systematic) literature review identified only eight studies reporting dispersal kernels for plant pathogens that used data gathered in experimental designs extending over regions bigger than 1 km² (Fabre et al., 2021). The mean dispersal distance of the four fungal pathosystems listed by these authors are 213 m for the ascospores of *Mycosphaerella fijiensis*, 490 m for the ascospores of *Leptosphaeria maculans*, 860 m for *Podosphaera plantaginis* and from 1380 to 2560 m for *Hymenoscyphus fraxineus*. Our estimates for poplar rust are in the same range as the one obtained for *Hymenoscyphus fraxineus*, the causal agent of Chalara ash dieback, at regional scale (Grosdidier et al., 2018).

Dispersal kernels can be further defined by their shape, which informs in particular the "fatness" of their tails. Here, we show that the spread of poplar rust is best described by a fat-tailed Exponential-power kernel whereas the "thin-tailed" kernels considered (Gaussian and exponential kernels) were clearly discarded by model selection. Even though we lacked a previous estimate of dispersal in *M. larici-populina*, these results are in accordance with the high dispersal ability and the long-distance dispersal events evidenced in this species from former population genetic analyses (Barrès et al., 2008; Becheler et al., 2016). Rust fungi are well-known to be wind dispersed over long distances (Brown and Hovmøller, 2002; Aylor, 2003), some studies explaining the spread of rust disease at the global scale through the air mass movement (Pan et al., 2006). Contrary to other wind-borne fungi like ascomycetes (which actively project the sexually derived spores, Roper et al., 2010), spores are released passively through air current, which results in both a high amount of auto-infection and long dispersal distances. Recently, Severns et al. (2019) gathered experimental and simulation evidence that support that wheat stripe rust spread supports theoretical scaling relationships from power law properties, another family of fat-tail dispersal kernel. Fat-tail kernels have also been identified in other plant pathogens. For example, Rieux et al. (2014) and Bousset et al. (2015) found evidence for Exponential-power kernels with fat-tailed for the causal agents of Black Sigatoka and Phoma stem canker, respectively. More broadly, four of these seven studies listed by Fabre et al. (2021) lent support to fat-tailed kernels, including plant pathogens as diverse as viruses, fungi and oomycetes. Many aerially dispersed pathogens are likely to display long-distance flights as soon as their life cycles include propagules. Typically, it can be spore propagules but also insect vectors such as aphids escaping from plant canopy into turbulent air layer (Ferrandino, 1993; Pan et al., 2010).

5.2 Effect of fat-tailed dispersal kernel on eco-evolutionary dynamics

The dynamics produced by the integro-differencial models we use strongly depends on the tail of the dispersal kernel. Namely, when the equation is homogeneous (*i.e.* when the model parameters did not vary in space, leading to $r(x) = r$), it is well known that for thin-tailed dispersal kernel J that is such that $\int_{\mathbb{R}} J(z)e^{\lambda|z|}dz < +\infty$ for some $\lambda > 0$, the dynamics of the solution $u(t, x)$ is well explained using a particular solution called travelling wave. Such particular solutions can then be used to show that the invading front describe by the solution $u(t, x)$ moves at a constant speed (Aronson and Weinberger, 1978). For a fat-tailed kernel, these particular solutions do not exist anymore and the dynamic of $u(t, x)$ describes an accelerated invasion process (Medlock and Kot, 2003; Garnier, 2011; Bouin et al., 2018) which can be quantified precisely. Here, the dispersal kernel and growth rates selected clearly suggest that the dynamics of the poplar rust could be better described as an accelerated invasion process rather than a front moving at a constant speed. Such accelerating wave at the epidemic front has been identified for several fungal plant pathogens dispersed by wind, including *Puccinia striiformis* and *Phytophthora infestans* the wheat stripe rust and the potato late blight, respectively (Mundt et al., 2009). However, it should be stated that fat-tailed kernels are not always associated with accelerated invasion processes. Indeed, fat-tailed kernels can be further distinguished depending on whether they are "regularly varying" (e.g. power law kernels) or "rapidly varying" (e.g. Exponential-power kernels) (Klein et al., 2006). Mathematically, it implies that power law kernels decrease even more slowly than any Exponential-power function. Biologically this means that fat-tailed Exponential-power kernels display rarer long-distance dispersal events than power law kernels. On the theoretical side, the kernel's properties (*i.e.* "regularly varying" and "rapidly varying" kernels) subtly interact with demographic mechanisms such as Allee effects to possibly cancel the acceleration of invasion. With weak Allee effects, *i.e.* the growth rate is density dependent but still positive, no acceleration can occur in the class of rapidly varying kernel whereas it could be observed for some regularly varying kernels depending on the strength of the density dependence (Alfaro and Coville, 2017; Bouin et al., 2021). For strong Allee effects (*i.e.* a negative growth rate at low density), no acceleration can be observed for all possible kernels (Chen, 1997). On the applied side, whether or not the epidemic wave is accelerating sharply impacts control strategies of emerging epidemics based on removing plants from fields contained within a given radius around the detected epidemic focus (Fabre et al., 2021). Similarly, taking into account the existence of long-distance dispersal is fundamental to design efficient management strategies, as shown for example for sudden oak death (Filipe et al., 2012) and for downy mildew of cucurbit (Ojiambo et al., 2015).

5.3 Confidence in the inference of the dispersal process

The inference framework we developed is reasonably efficient in estimating the dispersal process with frequent long-distance dispersal events as generated using integro-differential models with Exponential-power dispersal kernels. In particular, the numerical experiments clearly show that the lower the

parameter τ of Exponential-power dispersal kernels, the higher the confidence in the selection of this kernel. Conversely, the identification of the dispersal process is less accurate when data are simulated under the Gaussian kernel. Nonetheless, its correct identification can be highly improved by densifying the sampling. We clearly observed that integro-differential models with Gaussian dispersal kernel on the one hand and reaction-diffusion equation on the other hand are well identified with our estimation procedure when the time and space sampling is dense enough. This result may at first appear striking as a common belief tends to consider that diffusion amounts to a Gaussian dispersal kernel. However, these two models do not represent similar movement processes. Indeed, macroscopic diffusion based on stochastic diffusion motions does not model the possibility of long-distance dispersal events which is the specificity of integro-differential models. If it is reasonable to assume that an individual disperses via a Brownian motion (which amounts to diffusion) (Hapca et al., 2009), extending this assumption to every individual in the population does not seem appropriate given the variability among individuals' movements. Accordingly, we believe that integro-differential models are better suited to take this inter-individual variability into account. Indeed they explicitly model the redistribution of individuals according to a given dispersal kernel.

If one wants to distinguish which of diffusion or Gaussian kernel best represents the dispersal process of interest, we advocate densifying the sampling in order to properly differentiate them. Conversely, if the choice is between different fat-tailed dispersal kernels, there is no need to densify the sampling. Indeed, in these cases, the sampling densification will likely bring little additional information compared to its additional cost. Last, the dispersal estimation accuracy can be enhanced by improving the mechanistic model to take into account additional sources of variability (*e.g.* environmental parameters, weather, wind, etc.).

5.4 Robustness and portability of the method

A strength of the approach proposed is the detailed description of the observation laws in the statistical part of the model. First, the observation laws incorporate the sources of variability for the two distinct types of samples realised, as encapsulated in the parameters σ and γ. Second, the derivation of their probability density functions allows to obtain an analytical expression of the likelihood function and, as such, to rely on classical methods of maximum likelihood estimation. In turn, these optimization methods are a simpler and faster alternative to inferences relying for example on Bayesian approaches. However, the framework of hierarchical statistical models (Cressie et al., 2009), whose inference is often facilitated by Bayesian approaches, could likely be mobilised to improve model fit. In particular, although the coverage rate of the tree sampling was correct, it could be further improved by relaxing some hypotheses. In particular, orange-coloured uredinia being easily seen on green leaves, we assumed that the persons in charge of the sampling perfectly detect the disease as soon as a single uredinia is present on a leaf. However, even in this context, observation errors are likely present in our dataset as in any large spatio-temporal study. Hierarchical Bayesian models can be used to incorporate latent variables allowing to take into account that a tree observed to be healthy can actually be infected. Symmetrically, we could take into account false detection of infection. This would make

sense in our case study as a sister species, *M. alli-populina* can also infect poplar leaves. This species can predominate locally in the downstream part of the Durance River valley and is not easily discernible from *M. larici-populina* in the field. This could have led to over-estimate the disease severity at some locations. Yet, all infected leaves were collected and minutely inspected in the lab under a Stereo Microscope (25× magnification) to check for species identification.

More generally, the statistical part of the mechanistic-statistical approaches developed could be transposed to a wide range of organisms and sampling types. Here we consider the leaf as the sampling unit, and two distinct types of sampling: a sampling of random leaves in trees, and a sampling of leaves grouped within twigs. These types of sampling apply to a wide range of species, which local distribution is aggregated into patches randomly scattered across a study site. Any biological system with two such distinct types of sampling would fit the proposed statistical part of the model. One can for example scale up the sampling by considering the plant (instead of the leaf) as the basic unit. Moreover, our framework naturally copes with the diversity of sampling situations on the ground. In particular, it is possible to remove one of the two sample types considered, for all or part of the sampled sites and dates.

The mechanistic part of the model could also handle a wider diversity of hypotheses. First, the model can be adapted to take into account a wider range of dispersal kernels, typically regularly varying kernels (see above). Second, the model can also easily be adapted to take into account time and space discontinuities of its parameters. Typically, one may easily assume that the growth rate depends on daily meteorological variables. Finally, note that the model ignores the influence of the local fluctuations of the population size on the macro-scale density of the population. However, it has been shown that stochastic fluctuations can influence epidemic dynamics (for example by Rohani et al., 2002, using temporal epidemic models). Here, we neglect this influence by considering that the average population size is relevant when habitat units are aggregated. Relaxing this hypothesis could be achieved by incorporating stochastic integro-differential equations. The inference of such models is currently a front of research.

5.5 Future directions

Here we provide a sound inferential framework to infer parameters and estimate dispersal from count data when monitoring a colonisation event. During colonisation events, the Allee effect can be very important because it can modulate the invasion dynamics of certain species (Dennis, 1989; Lewis and Kareiva, 1993). An Allee effect could be introduced by modulating the population growth depending on its size, *via* a new parameter. It has been included in an earlier version of this model but led to practical identifiability problems between this Allee parameter and the population growth rates. Here we typically face a limitation caused by the type of data used. We argue that incorporating genetic data would allow to overcome this difficulty and properly infer the strength of the Allee effect.

As biological invasions are regularly observed retrospectively, carrying out spatiotemporal monitoring is often highly difficult, when possible. Not having data describing the temporal dimension of an invasion will make model inference very difficult, and much less accurate. It would be difficult to

properly disentangle the effect of growth rate and dispersal. Here again, gathering genetic data may be relevant as colonisation and demographic effects generate their own specific genetic signatures (Miller et al., 2020). Furthermore, genetic data could help to identify the dispersal kernel underlying the invasion process. Indeed, the population will exhibit an erosion of its neutral diversity with a thin-tailed kernel (Edmonds et al., 2004; Hallatschek et al., 2007). This fact can be apprehended by an analysis the behaviour of neutral fraction $u_i(t,x)$ of the solution $u(t,x)$, ($u(t,x) = \sum_{i\in\mathcal{I}} u_i(t,x)$, where $\mathcal{I}$ is a set of genotypes). The composition of the invasion front can then be analytically studied. For thin-tailed kernels, the invasion front described by $u(t,x)$ is progressively mostly composed of the neutral fraction that spreads at the forefront of the colonisation wave. Conversely, genetic diversity can be preserved all along the invasion front with a fat-tailed kernel, because of the long-distance dispersal of individuals from the back of the front, where genetic diversity is conserved (Fayard et al., 2009; Bonnefon et al., 2014).

Acknowledgements

This work was supported by grants from the French National Research Agency (ANR-18-CE32-0001, CLONIX2D project; ANR-11-LABX-0002-01, Cluster of Excellence ARBRE). Méline Saubin was supported by a PhD fellowship from INRAE and the French National Research Agency (ANR-18-CE32-0001, CLONIX2D project).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alfaro, M. and Coville, J. (2017). Propagation phenomena in monostable integro-differential equations: Acceleration or not? *Journal of Differential Equations*, 263(9):5727–5758.

Aronson, D. G. and Weinberger, H. F. (1978). Multidimensional nonlinear diffusion arising in population genetics. *Advances in Mathematics*, 30(1):33–76.

Aylor, D. E. (2003). Spread of plant disease on a continental scale: Role of aerial dispersal of pathogens. *Ecology*, 84(8):1989–1997.

Barrès, B., Halkett, F., Dutech, C., Andrieux, A., Pinon, J., and Frey, P. (2008). Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, 8(5):577–587.

Becheler, R., Xhaard, C., Klein, E., Hayden, K. J., Frey, P., De Mita, S., and Halkett, F. (2016). Genetic signatures of a range expansion in natura: when clones play leapfrog. *Ecology and Evolution*, 6(18):6625–6632.

Berliner, L. M. (2003). Physical-statistical modeling in geophysics. *Journal of Geophysical Research*, 108(24):8776.

Bialozyt, R., Ziegenhagen, B., and Petit, R. J. (2006). Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*, 19(1):12–20.

Bonnefon, O., Coville, J., Garnier, J., Hamel, F., and Roques, L. (2014). The spatio-temporal dynamics of neutral genetic diversity. *Ecological Complexity*, 20:282–292.

Bouin, E., Coville, J., and Legendre, G. (2021). Sharp exponent of acceleration in general nonlocal equations with a weak Allee effect. *arXiv*, pages 1–45.

Bouin, E., Garnier, J., Henderson, C., and Patout, F. (2018). Thin front limit of an integro-differential Fisher-KPP equation with fat-tailed kernels. *SIAM Journal on Mathematical Analysis*, 50(3):3365–3394.

Bousset, L., Jumel, S., Garreta, V., Picault, H., and Soubeyrand, S. (2015). Transmission of *Leptosphaeria maculans* from a cropping season to the following one. *Annals of applied biology*, 166(3):530–543.

Broquet, T. and Petit, E. J. (2009). Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):193–216.

Brown, J. K. M. and Hovmøller, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science's compass*, 297:537–541.

Chagneau, P., Mortier, F., Picard, N., and Bacro, J. (2011). A hierarchical Bayesian model for spatial prediction of multivariate non-Gaussian random fields. *Biometrics*, 67(1):97–105.

Chen, X. (1997). Existence, uniqueness, and asymptotic stability of traveling waves in nonlocal evolution equations. *Advances in Differential Equations*, 2(1):125–160.

Clark, J. S., Lewis, M., and Horvath, L. (2001). Invasion by extremes: Population spread with variation in dispersal and reproduction. *American Naturalist*, 157(5):537–554.

Clobert, J., Ims, R. A., and Rousset, F. (2004). Causes, mechanisms and consequences of dispersal. In *Ecology, genetics and evolution of metapopulations*, pages 307–335. Elsevier.

Cressie, N., Calder, C. A., Clark, J. S., Ver Hoef, J. M., and Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, 19(3):553–570.

Dennis, B. (1989). Allee effects: Population growth, critical density, and the chance of extinction. *Natural Resource Modeling*, 3(4):481–538.

Dybiec, B., Kleczkowski, A., and Gilligan, C. A. (2009). Modelling control of epidemics spreading by long-range interactions. *Journal of the Royal Society Interface*, 6(39):941–950.

Edmonds, C. A., Lillie, A. S., and Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences*, 101(4):975–979.

Fabre, F., Coville, J., and Cunniffe, N. J. (2021). Optimising reactive disease management using spatially explicit models at the landscape scale. In Scott, P. R., Strange, R. N., Korsten, L., and Gullino, M. L., editors, *Plant disease and food security in the 21st century*, pages 47–72. Springer International Publishing.

Fayard, J., Klein, E., and Lefèvre, F. (2009). Long distance dispersal and the fate of a gene from the colonization front. *Journal of Evolutionary Biology*, 22(11):2171–2182.

Ferrandino, F. J. (1993). Dispersive epidemic waves: I. Focus expansion within a linear planting. *Phytopathology*, 83(8):795.

Ferrandino, F. J. (1996). Lenght scale of disease spread: Fact or artifact of experimental geometry. *Phytopathology*, 86:806–811.

Filipe, J. A. N., Cobb, R. C., Meentemeyer, R. K., Lee, C. A., Valachovic, Y. S., Cook, A. R., Rizzo, D. M., and Gilligan, C. A. (2012). Landscape epidemiology and control of pathogens with cryptic and long-distance dispersal: Sudden Oak death in northern Californian forests. *PLoS Computational Biology*, 8(1):e1002328.

Fisher, R. A. (1937). The wave of advance of advantageous genes. *Annals of Eugenics*, 7:355–369.

Garnier, J. (2011). Accelerating solutions in integro-differential equations. *SIAM J. Math. Anal.*, 43:1955–1974.

Georgescu, V., Desassis, N., Soubeyrand, S., Kretzschmar, A., and Senoussi, R. (2014). An automated MCEM algorithm for hierarchical models with multivariate and multitype response variables. *Communications in Statistics - Theory and Methods*, 43(17):3698–3719.

Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648.

Grosdidier, M., Ioos, R., Husson, C., Cael, O., Scordia, T., and Marçais, B. (2018). Tracking the invasion: dispersal of *Hymenoscyphus fraxineus* airborne inoculum at different scales. *FEMS microbiology ecology*, 94(5):1–11.

Hallatschek, O. and Fisher, D. S. (2014). Acceleration of evolutionary spread by long-range dispersal. *Proceedings of the National Academy of Sciences*, 111(46):E4911–E4919.

Hallatschek, O., Hersen, P., Ramanathan, S., and Nelson, D. R. (2007). Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930.

Hapca, S., Crawford, J. W., and Young, I. M. (2009). Anomalous diffusion of heterogeneous populations characterized by normal diffusion at the individual level. *Journal of the Royal Society Interface*, 6(30):111–122.

Hefley, T. J., Hooten, M. B., Russell, R. E., Walsh, D. P., and Powell, J. A. (2017). When mechanism matters: Bayesian forecasting using models of ecological diffusion. *Ecology Letters*, 20(5):640–650.

Hutson, V., Martinez, S., Mischaikow, K., and Vickers, G. T. (2003). The evolution of dispersal. *Journal of Mathematical Biology*, 47(6):483–517.

Ibrahim, K. M., Nichols, R. A., and Hewitt, G. M. (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, 77:282–291.

Klein, E., Lavigne, C., Picault, H., Renard, M., and Gouyon, P. H. (2006). Pollen dispersal of oilseed rape: Estimation of the dispersal function and effects of field dimension. *Journal of Applied Ecology*, 43(1):141–151.

Kolmogorov, A. N., Petrovsky, I. G., and Piskunov, N. S. (1937). Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Bulletin Université d'État à Moscow (Bjul. Moskowskogo Gos. Univ)*, pages 1–26.

Kot, M., Lewis, M. A., and van den Driessche, P. (1996). Dispersal data and the spread of invading organisms. *Ecology*, 77(7):2027–2042.

Kuparinen, A., Snäll, T., Vänskä, S., and O'Hara, R. B. (2007). The role of model selection in describing stochastic ecological processes. *Oikos*, 116(6):966–974.

Lewis, M. A. and Kareiva, P. (1993). Allee dynamics and the spread of invading organisms. *Theoretical Population Biology*, 42:141–158.

Louvrier, J., Papaïx, J., Duchamp, C., and Gimenez, O. (2020). A mechanistic–statistical species distribution model to explain and forecast wolf (*Canis lupus*) colonization in South-Eastern France. *Spatial Statistics*, 36:100428.

Maupetit, A., Larbat, R., Pernaci, M., Andrieux, A., Guinet, C., Boutigny, A. L., Fabre, B., Frey, P., and Halkett, F. (2018). Defense compounds rather than nutrient availability shape aggressiveness trait variation along a leaf maturity gradient in a biotrophic plant pathogen. *Frontiers in Plant Science*, 9(September).

Medlock, J. and Kot, M. (2003). Spreading disease: Integro-differential equations old and new. *Math. Biosci.*, 184(2):201–222.

Miller, T. E. X., Angert, A. L., Brown, C. D., Lee-Yaw, J. A., Lewis, M., Lutscher, F., Marculis, N. G., Melbourne, B. A., Shaw, A. K., Szűcs, M., Tabares, O., Usui, T., Weiss-Lehman, C., and Williams, J. L. (2020). Eco-evolutionary dynamics of range expansion. *Ecology*, 101(10):1–14.

Mollison, D. (1977). Spatial contact models for ecological and epidemic spread. *J. R. Stat. Ser. B Stat. Methodol.*, 39:283–326.

Mundt, C. C., Sackett, K. E., Wallace, L. D., Cowger, C., and Dudley, J. P. (2009). Long-distance dispersal and accelerating waves of disease: Empirical relationships. *American Naturalist*, 173(4):456–466.

Murray, J. D. (2002). *Mathematical Biology*, volume 17. Springer-Verlag, third edition.

Nathan, R. (2001). The challenges of studying dispersal. *Trends in Ecology and Evolution*, 16(9):481–483.

Nathan, R., Klein, E., Robledo-Arnuncio, J., and Revilla, E. (2012). Dispersal kernel: Review. *Oxford University Press*.

Nathan, R., Perry, G., Cronin, J. T., Strand, A. E., and Cain, M. L. (2003). Methods for estimating long-distance dispersal. *Oikos*, 103(2):261–273.

Nichols, R. A. and Hewitt, G. M. (1994). The genetic consequences of long distance dispersal during colonization. *Heredity*, 72:312–317.

Ojiambo, P. S., Gent, D. H., Quesada-Ocampo, L. M., Hausbeck, M. K., and Holmes, G. J. (2015). Epidemiology and population biology of Pseudoperonospora cubensis : A model system for management of downy mildews. *Annual Review of Phytopathology*, 53(1):223–246.

Okubo, A. and Levin, S. A. (2002). *Diffusion and Ecological Problems – Modern Perspectives*. Second edition, Springer-Verlag, New York.

Pan, Z., Li, X., Yang, X. B., Andrade, D., Xue, L., and McKinney, N. (2010). Prediction of plant diseases through modelling and monitoring airborne pathogen dispersal. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, 5(018).

Pan, Z., Yang, X. B., Pivonia, S., Xue, L., Pasken, R., and Roads, J. (2006). Long-term prediction of soybean rust entry into the continental United States. *Plant Disease*, 90(7):840–846.

Papaïx, J., Soubeyrand, S., Bonnefon, O., Walker, E., Louvrier, J., Klein, E., and Roques, L. (2022). Inferring mechanistic models in spatial ecology using a mechanistic-statistical approach. In *Statistical Approaches for Hidden Variables in Ecology*, pages 69–95. Wiley.

Petit, R. J. (2004). Biological invasions at the gene level. *Diversity and Distributions*, 10(3):159–165.

R (2018). R Core Team, R: A language and environment for statistical computing.

Rieux, A., Lenormand, T., Carlier, J., De Lapeyre de Bellaire, L., and Ravigné, V. (2013). Using neutral cline decay to estimate contemporary dispersal: A generic tool and its application to a major crop pathogen. *Ecology Letters*, 16(6):721–730.

Rieux, A., Soubeyrand, S., Bonnot, F., Klein, E. K., Ngando, J. E., Mehl, A., Ravigne, V., Carlier, J., and Bellaire, L. (2014). Long-distance wind-dispersal of spores in a fungal plant pathogen: Estimation of anisotropic dispersal kernels from an extensive field experiment. *PLoS ONE*, 9(8):e103225.

Ritz, C., Baty, F., Streibig, J. C., and Gerhard, D. (2015). Dose-Response analysis using R. *PLOS ONE*, 10(12).

Rohani, P., Keeling, M. J., and Grenfell, B. T. (2002). The interplay between determinism and stochasticity in childhood diseases. *The American Naturalist*, 159:469–481.

Roper, M., Seminara, A., Bandi, M. M., Cobb, A., Dillard, H. R., and Pringle, A. (2010). Dispersal of fungal spores on a cooperatively generated wind. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41):17474–17479.

Roques, L., Soubeyrand, S., and Rousselet, J. (2011). A statistical-reaction-diffusion approach for analyzing expansion processes. *Journal of Theoretical Biology*, 274(1):43–51.

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., Mcroberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature ecology evolution*, 3(3):430–439.

Severns, P. M., Sackett, K. E., Farber, D. H., and Mundt, C. C. (2019). Consequences of long-distance dispersal for epidemic spread: Patterns, scaling, and mitigation. *Plant Disease*, 103(2):177–191.

Shigesada, N. and Kawasaki, K. (1997). *Biological invasions: Theory and practice*. Oxford University Press, UK.

Soubeyrand, S., Neuvonen, S., and Penttinen, A. (2009). Mechanical-statistical modeling in ecology: From outbreak detections to pest dynamics. *Bull. Math. Bio.*, 71:318–338.

Soubeyrand, S. and Roques, L. (2014). Parameter estimation for reaction-diffusion models of biological invasions. *Population Ecology*, 56(2):427–434.

Strange, R. N. and Scott, P. R. (2005). Plant disease: A threat to global food security. *Annual Review of Phytopathology*, 43:83–116.

Szymańska, Z., Skrzeczkowski, J., Miasojedow, B., and Gwiazda, P. (2021). Bayesian inference of a non-local proliferation model. *Royal Society Open Science*, 8(11).

Wikle, C. K. (2003). Hierarchical models in environmental science. *International Statistical Review*, 71:181–199.

Xhaard, C., Barrès, B., Andrieux, A., Bousset, L., Halkett, F., and Frey, P. (2012). Disentangling the genetic origins of a plant pathogen during disease spread using an original molecular epidemiology approach. *Molecular Ecology*, 21(10):2383–2398.

Appendix

A Numerical scheme

We use an implicit Euler scheme combined with a finite difference scheme to compute the solution $u(t,x)$ of the reaction-diffusion equation over $[-R,R]\times[0,T]$, with $2\times R$ the length of the modelled environment, and T the duration of the modelled process. For the integro-differential equation, we use an explicit Euler scheme. More precisely, we perform a standard explicit Euler time discretisation of the equation:

$$\frac{\partial u}{\partial t}(t,x) \approx \frac{u(t+\delta,x)-u(t,x)}{\delta},$$

that leads to:

$$\begin{aligned} u(t_{n+1},x) = u(t_n,x) &+ \delta\left(\int_{-R}^{R} J(x-y)[u(t_n,y)-u(t_n,x)]\,dy\right) \\ &+ \delta r(x)u(t_n,x)\left(1-\frac{u(t_n,x)}{K}\right), \end{aligned} \tag{S1}$$

where $\{t_n = n\delta = nT/N : n = 0,\dots,N\}$ is a series of increasing times separated by $\delta = T/N > 0$, and N is the number of time steps in the series. For the space discretisation, we define a regular grid $\{x_i = -R + i\epsilon = -R + 2Ri/I : i = 0,\dots,I\}$ with $I+1$ points separated by $\epsilon = 2R/I > 0$. We make the following approximation for all x in $[-R,R]$:

$$u(t_n,x) \approx \sum_{i=0}^{I} u(t_n,x_i)\mathbb{1}_{[x_i,x_i+\epsilon)}(x), \tag{S2}$$

where $x \mapsto \mathbb{1}_{[x_i,x_i+\epsilon)}(x)$ is the indicator function that gives 1 if $x\in[x_i,x_i+\epsilon)$, 0 otherwise. Based on this approximation, we only need to compute $u(t,x)$ at points x_i, $i=0,\dots,I$. Plugging Approximation (S2) in the integral of Equation (S1) computed for $x=x_i$ yields:

$$\begin{aligned} &\int_{-R}^{R} J(x_i-y)[u(t_n,y)-u(t_n,x_i)]\,dy \\ &\qquad\approx \int_{-R}^{R} J(x_i-y)\left[\left(\sum_{j=0}^{I} u(t_n,x_j)\mathbb{1}_{[x_j,x_j+\epsilon)}(y)\right) - u(t_n,x_i)\right] dy \\ &\qquad= \left(\sum_{j=0}^{I} u(t_n,x_j)\int_{-R}^{R} J(x_i-y)\mathbb{1}_{[x_j,x_j+\epsilon]}(y)\,dy\right) - \left(u(t_n,x_i)\int_{-R}^{R} J(x_i-y)dy\right) \\ &\qquad\approx \epsilon\left(\sum_{j=0}^{I} u(t_n,x_j)J(x_i-x_j)\right) - \epsilon\, u(t_n,x_i)\sum_{j=0}^{I} J(x_i-x_j). \end{aligned}$$

Let us define the matrix $\mathbf{J}^{in} := (J(x_i - x_j))_{0 \leq i,j \leq I}$ whose element (i,j) is $\mathbf{J}^{in}_{ij} = J(x_i - x_j)$. We get the following numerical scheme:

$$u(t_{n+1}, x_i) = u(t_n, x_i) + \delta\epsilon \left[\sum_{j=0}^{I} \mathbf{J}^{in}_{ij} u(t_n, x_j) - u(t_n, x_i) \left(\sum_{j=0}^{I} \mathbf{J}^{in}_{ij} \right) \right] + \delta r(x_i) u(t_n, x_i) \left[1 - \frac{u(t_n, x_i)}{K} \right].$$

By defining the vectors $\mathbf{U}(t_n) = (u(t_n, x_i))_{0 \leq i \leq I}$, $\mathbf{R} = (r(x_i))_{0 \leq i \leq I}$ and $\mathbf{1} = (1)_{0 \leq i \leq I}$, we have to solve the linear system:

$$\mathbf{U}(t_{n+1}) = \mathbf{U}(t_n) + \delta\epsilon \left\{ \mathbf{J}^{in} \mathbf{U}(t_n) - \mathbf{U}(t_n) \cdot (\mathbf{J}^{in} \mathbf{1}) \right\} + \delta \{ \mathbf{R} \cdot \mathbf{U}(t_n) \} \cdot \left\{ \left(1 - \frac{\mathbf{U}(t_n)}{K} \right) \right\}.$$

where $\cdot$ is the element-wise multiplication operator.

B Distributions of the population measurements

B.1 Term designations for the sampling units

In our biological application, a poplar leaf represents a habitat unit, a twig represents a group of habitat units, and a tree represents a habitat bloc. For clarity, we refer to leaves, twigs and trees in the following explanations. We call a sampling site a surveyed area along the valley, containing several hundreds of trees. Further adaptations of this model to other sampling units would only require adapting this initial vocabulary.

B.2 Raw sampling

In the raw sampling, trees represent the sampling units, and B_{st} trees are observed in site s at time t. For each tree $b \in \{1, \ldots, B_{st}\}$, we measure the presence/absence of the pathogen by monitoring an equivalent number of M leaves within b (see Appendix C below for the determination of M). A tree is infected if at least one pathogen lesion has been detected, in at least one leaf of the tree. The observation in site s at time t is the number Y_{st} of infected trees.

Now, let us derive the probabilistic law of the presence/absence of the pathogen in any tree b observed in site s at time t. In this paragraph, subscripts s, t, and b are generally omitted to avoid cumbersome notation. We first remind that the numbers of pathogen lesions $N_i(t)$ in the leaf $i \in \{1, \ldots, M\}$ observed in tree b, given $R_i(t)$ and $u(t, x_s)$, are independent and Poisson distributed (see Eq. (1) in the main text):

$$N_i(t) \mid u(t, x_s), R_i(t) \sim_{indep.} \text{Poisson}(u(t, x_s) R_i(t)). \tag{S3}$$

In the raw sampling, M leaves are sampled at different locations on the tree (*i.e.* they belong to different groups, referred to as twigs), but further information about the twigs is not known. Thus, in the following, we take into account the twig structure without exploiting twig information. The leaves of a

given twig g on tree b share at time t the same suitability $\mathcal{R}_g(t)$, which is unobserved and Gamma distributed like in Eq. (2) in the main text (for all leaves i in twig g, $R_i(t) = \mathcal{R}_g(t)$). Given the suitabilities $\{\mathcal{R}_g(t) : g = 1, \ldots, G\}$ of twigs which compose tree b and given the absence of data about the twigs, $R_i(t)$ ($i \in \{1, \ldots, M\}$) are independent and identically distributed under the discrete empirical probability distribution:

$$\hat{F}_G(r) = \frac{1}{G} \sum_{g=1}^{G} \mathbb{1}(r \leq \mathcal{R}_g(t)), \tag{S4}$$

where $\mathbb{1}(\cdot)$ is the indicator function. Therefore, $N_i(t)$ ($i \in \{1, \ldots, M\}$) given $\{\mathcal{R}_g(t) : g = 1, \ldots, G\}$ and $u(t, x_s)$ are independent and their probability distribution is, using Eqs. (S3)–(S4):

$$P[N_i(t) = n \mid u(t, x_s), \{\mathcal{R}_g(t) : g = 1, \ldots, G\}] = \frac{1}{G} \sum_{g=1}^{G} \exp(-u(t, x_s)\mathcal{R}_g(t)) \frac{(-u(t, x_s)\mathcal{R}_g(t))^n}{n!}. \tag{S5}$$

The suitability $\mathcal{R}_g(t)$ being Gamma distributed with shape and scale parameters σ^{-2} and σ^2, respectively, the right-hand-side of Eq. (S5) is a Monte Carlo approximation of the integral:

$$\begin{aligned} &\int_{\mathbb{R}_+} \exp(-u(t, x_s)r) \frac{(-u(t, x_s)r)^n}{n!} \frac{1}{(\sigma^2)^{\sigma^{-2}}\Gamma(\sigma^{-2})} r^{\sigma^{-2}-1} e^{-r/\sigma^2} dr \\ &\quad = \frac{\Gamma(n + \sigma^{-2})}{(n!)\Gamma(\sigma^{-2})} \left(1 - \frac{u(t, x_s)}{u(t, x_s) + \sigma^{-2}}\right)^{\sigma^{-2}} \left(\frac{u(t, x_s)}{u(t, x_s) + \sigma^{-2}}\right)^n, \end{aligned} \tag{S6}$$

which coincides with the probability distribution of the negative–Binomial law (*i.e.* the Gamma-Poisson mixture distribution) given by Eq. (3) in the main text. The larger G, the more precise the approximation. Consequently, $N_i(t)$ ($i \in \{1, \ldots, M\}$) given $u(t, x_s)$ are asymptotically independent and distributed under the negative-Binomial distribution given by Eq. (3) in the main text. Based on this approximation, the infections of leaves from tree b in site s at time t are asymptotically independent and distributed under Bernoulli distributions with success probability:

$$\begin{aligned} p_{st}^{\text{leaf}} &= P(N_i(t) > 0 \mid u(t, x_s)) \\ &= 1 - P(N_i(t) = 0 \mid u(t, x_s)) \\ &= 1 - (1 + u(t, x_s))^{-1/\sigma^2}, \end{aligned} \tag{S7}$$

The people who carried out the sampling observed a number M of leaves on tree b. Due to the particular configuration of the foliage of each tree, we assumed that the number Y_{stb}^{leaf} of infected leaves among the M leaves observed in tree b is approximately distributed under a Beta-Binomial distribution with mean Mp_{st}^{leaf} and tree perception parameter γ :

$$Y_{stb}^{\text{leaf}} \mid u(t, x_s) \sim_{approx.} \text{Beta-Binomial}(M, p_{st}^{\text{leaf}}, \gamma). \tag{S8}$$

Accordingly, the probability, as *perceived* by people in charge of the sampling, of leaf infection on the set of M leaves observed on a given tree, is distributed according to a Beta distribution. The Beta distribution is centred around the true probability of leaf infection p_{st}^{leaf} and allows *perceived* probability to vary from tree to tree depending on the tree perception parameter γ. It follows that the infection of

tree b is approximately distributed under the Bernoulli distribution with success probability:

$$\begin{aligned} p_{st}^{\text{tree}} &= P(Y_{stb}^{\text{leaf}} > 0 \mid u(t, x_s)) \\ &= 1 - P(Y_{stb}^{\text{leaf}} = 0 \mid u(t, x_s)) \\ &= 1 - \frac{Beta[\gamma p_{st}^{\text{leaf}}, M + \gamma(1 - p_{st}^{\text{leaf}})]}{Beta[\gamma p_{st}^{\text{leaf}}, \gamma(1 - p_{st}^{\text{leaf}})]} \end{aligned} \tag{S9}$$

where p_{st}^{leaf} is given by S7 and $Beta$ represents the beta function. It follows that the probability distribution functions of the number Y_{st}^{tree} of infected trees infected among the B_{st} trees observed satisfy, for all sampling sites s and sampling times t

$$\begin{aligned} f_{st}^{raw}(y) &= P[Y_{st}^{\text{tree}} = y \mid u(t, x_s)] \\ &= f_{Binomial(B_{st}, p_{st,M}^{\text{tree}})}(y) \end{aligned} \tag{S10}$$

where $f_{Binomial}$ is the density of the Binomial distribution.

B.3 Refined sampling

In the refined sampling, G_{st} twigs (*i.e.* groups of spatially connected leaves) are sampled in site s at time t. Here, the twig information (the number of twigs and the distribution of leaves on twigs) are known but the suitability $\mathcal{R}_g(t)$ of leaves in a twig g remains unobserved. The numbers of pathogen lesions $N_i(t)$ in the observed leaves $i \in \{1, \ldots, M_{stg}\}$ of twig g given $\mathcal{R}_g(t)$ and $u(t, x_s)$ are independent and Poisson distributed:

$$N_i(t) \mid u(t, x_s), \mathcal{R}_g(t) \sim_{indep.} \text{Poisson}(u(t, x_s)\mathcal{R}_g(t)). \tag{S11}$$

Then, the numbers of infected leaves Y_{stg}^{leaf} (*i.e.* leaves with at least one pathogen lesion) given $\mathcal{R}_g(t)$ and $u(t, x_s)$ are independent and distributed under the following Binomial distributions:

$$Y_{stg}^{\text{leaf}} \mid u(t, x_s), \mathcal{R}_g(t) \sim_{\text{indep.}} \text{Binomial}(M_{stg}, 1 - e^{-u(t,x_s)\mathcal{R}_g(t)}). \tag{S12}$$

In addition,

$$u(t, x_s)\mathcal{R}_g(t) \mid u(t, x_s) \sim_{\text{indep.}} \text{Gamma}(\sigma^{-2}, u(t, x_s)\sigma^2). \tag{S13}$$

Using Eqs. (S11)–(S13), Y_{stg}^{leaf} given $u(t, x_s)$ are independent and follow Gamma-Binomial mixture distributions:

$$\begin{aligned} f_{st}^{ref}(y) &= P[Y_{stg}^{\text{leaf}} = y \mid u(t, x_s)] \\ &= \int_0^\infty f_{\text{Binomial}(M_{stg}, 1-e^{-z})}(y) f_{\text{Gamma}(\sigma^{-2}, u(t,x_s)\sigma^2)}(z) dz \end{aligned} \tag{S14}$$

where f_{Gamma} is the density of the Gamma distribution. Note that this Gamma-Binomial mixture

distribution is an over-dispersed Binomial distribution like the Beta-Binomial distribution.

C Estimation of the number of leaves efficiently observed during tree scans

A problem inherent to the raw sampling design is that we do not know the number of leaves observed during the scan of the trees, contrary to the twig data for which we counted both the number of infected leaves and the total number of leaves carried by each observed twig. In other words, an inspected tree is a set of leaves of unknown size.

Let us recall, however, that Eq. (S9) gives at each site s and date t the probability p_{st}^{tree} that a tree is infected as a function of both the probability p_{st}^{leaf} that a leaf is infected and the number M of leaves observed on a tree:

$$p_{st}^{\text{tree}} = 1 - (1 - p_{st}^{\text{leaf}})^M$$

Thus, the number of leaves on a tree satisfies:

$$M = \frac{\log(1 - p_{st}^{\text{tree}})}{\log(1 - p_{st}^{\text{leaf}})}$$

Let us use as approximations of p_{st}^{tree} the observed proportions q_{st}^{tree} of infected trees at sites s and dates t, and as approximations of p_{st}^{leaf} the observed proportions q_{st}^{leaf} of infected leaves (calculated from twig data). Then, an estimate $\hat{\lambda}_M$ of the mean number of leaves λ_M by tree is given by:

$$\hat{\lambda}_M = \text{round}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\log(1 - q_{st}^{\text{tree}})}{\log(1 - q_{st}^{\text{leaf}})}\right),$$

with N the number of pairs (s, t) (*i.e.* sampling sites and dates) displaying both tree and twig data. Proportions of infection $q_{st}^{\text{tree}} = 1$ and $q_{st}^{\text{leaf}} = 1$ where approximated to $1-10^{-16}$ for numerical considerations. This procedure led to $\hat{\lambda}_M = 10$.

This value may appear low. However, λ_M does not correspond to the actual mean number of leaves carried by an entire young tree but amounts to the mean number of leaves effectively inspected during tree scan, *i.e.* those observed as minutely as for the twig data in a limited time. It is important to note that for each tree the tree scan stops when an infected leaf is observed, or after 30 s of inspection. Therefore, the number of inspected leaves per tree can be very low in highly infected sites.

For the practical identifiability studies, we set $\lambda_M = 10$. For parameters inference on the real data set a different value of $(\hat{\lambda}_M)_t$ was estimated for each sampling date, from the observed proportions q_{st}^{tree} of infected trees and the observed proportions q_{st}^{leaf} of infected leaves at date t (Table S1).

Table S1: For parameters inference on the real data set, estimated number of leaves effectively observed per tree ($\hat{\lambda}_M$) for each sampling date.

Date t	$(\hat{\lambda}_M)_t$
1	40
2	24
3	6
4	3
5	5
6	1

D Simulation details

Computations were performed with the R software environment (R, 2018). The initial vector of population densities $u(0, x)$ for x over $[-R, R]$ was estimated from the data of the first sampling date, by fitting a general model for analysis of dose-response data (package DRC on R, Ritz et al., 2015). This initial vector of population densities represented the initial condition of all simulations. We modelled $N = 1500$ time steps and $I = 400$ points in space. Because of the numerical scheme, with these parameters the reaction-diffusion dispersal model RD required an upper limit for parameter λ: we set $\lambda_{up} = 23$ for this model.

To fit our real case study, for all simulations we set $R = 100$ km, for a 200 km long river valley, and the epidemic was monitored over $T = 150$ days. We considered a shift in the environment topology at $d = 0.31\%$ of the valley, which corresponds to the delimitation observed in the Durance River valley with the Serre-Ponçon dam at 62 km downstream of the starting point of the epidemic. Therefore, for all simulations, the two growth rates r_{up} and r_{dw} apply to continuous segments of proportions d and $1 - d$ of the monitored space, respectively.

D.1 Practical identifiability

Simulations were performed as follows in three steps.

Step 1 : Simulation of a realistic epidemic. Given a hypothetical dispersal model (J_{Exp}, J_{Gauss}, J_{ExpP} or RD), a candidate vector of parameter's values $\theta = (\theta_r, \theta_J, \gamma, \sigma^2)$ is independently and randomly drawn from dedicated distributions that encompass a large diversity of invading scenarios (Table S2). We then simulate the corresponding epidemic. This epidemic is considered 'realistic' if a set of requirements on the observed proportion of infected trees $P_{s,t}$ on the farther downstream site ($s = 2R$) is met:

- $P_{2R,30} < 0.1$ (the proportion of infected trees after one month is lower than 10%)
- $P_{2R,75} < 0.5$ (the proportion of infected trees after two and a half months is lower than 50%)
- $P_{2R,150} > 0.1$ (the proportion of infected trees after five months is higher than 10%)
- $P_{2R,150} < 0.8$ (the proportion of infected trees after five months is lower than 80%)

Step 1 is repeated until a candidate vector θ satisfies all the conditions required for a so-called realistic epidemic. This vector is denoted θ_{true} thereafter.

Table S2: Distribution used to randomly sample the vector of model's parameters θ.

Parameter	Distribution	Interval
r_{dw}	Log-Uniform	[0.01, 0.5]
ω	Uniform	[-2, 3]
λ	Log-Uniform	[0.2,5]
τ	Log-Uniform	[0.2,1]
γ	Log-Uniform	[2,20]
σ^2	Log-Uniform	[0.01, 15]

Step 2 : Simulation of the sampling process. We consider a sampling design similar to our real experiment with 6 sampling dates with 12 locations regularly spread over 150 days and 200 km ($R = 100$). As for our real data, we increase the location density for the fifth date, with 45 locations instead of 12. For each date and location, a raw sampling with 100 trees of 10 leaves each and a refined sampling with 20 groups of 25 spatially connected leaves each are simulated given θ_{true}. The resulting dataset is denoted $\mathcal{D}_{true}$.

Step 3 : Parameter's estimation. We use the data $\mathcal{D}_{true}$ to re-estimate the model's parameters by minimizing the logarithm of the likelihood function $L(\theta)$. We proceed in three substeps to find initial values giving a satisfactory rate of convergence.

Step 3.1 : Using step 1, we generate 500 vectors θ_{init}. Note that this step was only performed once for all the estimations performed in this article.

Step 3.2 : The corresponding 500 likelihood values $L(\theta_{init})$ are calculated given $\mathcal{D}_{true}$ and then order by increasing values. The first 20 vectors θ_{init} (corresponding to the lowest likelihoods) are then used as initial values for 50 steps of a NELDER-MEAD optimisation routines (R function `optim`), resulting in the estimation of 20 vectors θ_{init2}. The estimated θ_{init2} that do not satisfy lower bounds θ_{low} and upper bounds θ_{up} are excluded. We used $\theta_{low} = (r_{dw} = 0.001, \omega = -7, \lambda = 0.02, \tau = 0.02, \gamma = 1.05, \sigma^2 = 10^{-7})$ and $\theta_{up} = (r_{dw} = 0.5, \omega = 3, \lambda = 10, \tau = 1, \gamma = 30, \sigma^2 = 20)$, with $\lambda_{up} = 23$ instead of 10 for dispersal RD. The likelihood values of the n_{init} remaining vectors $L(\theta_{init2})$ are calculated (given $\mathcal{D}_{true}$) and, again, order by increasing likelihood.

Step 3.3 : θ is then estimated using NLMINB optimisation routines with lower and upper bounds θ_{low} and θ_{up}, respectively. The initial condition is set to the first vector θ_{init2} as ordered in the previous step. The estimated parameter's values θ_{estim} are accepted if the `nlminb` function delivered a successful convergence diagnostic (with tunning parameters rel.tol=5.10^{-5} and iter.max=3000). If not, the second vector θ_{init2} is used, and so on until reaching convergence or testing the n_{init} initial vector's values selected at step 3.2. In the latter case, a convergence failure is obtained. Overall, this algorithm allows to obtain high rates of convergence.

These three steps were reiterated until deriving the estimation of $n = 100$ realistic epidemics for each dispersal model. Checking for parameter's practical identifiability basically relies on plotting for each dispersal model the correlation between θ_{true} and θ_{estim} (Figures S1, S2, S3, S4). Among all simulations performed, the proportions of convergence were 0.91, 0.95, 0.93, and 0.90 for dispersal J_{Exp}, J_{Gauss}, J_{ExpP}, and RD, respectively. A simulation converged when the convergence diagnostic of the algorithm indicated a convergence, and when all parameters were estimated inside intervals of realistic values. In the small number of simulations where the value of λ_{estim} proposed by the optimisation algorithm was higher than 23 (which is the upper limit of our numerical scheme (Appendix D)), the simulation was still considered convergent with $\lambda_{estim} = 23$. This configuration can occur in particular when trying to fit dispersal RD on datasets simulated according to (J_{ExpP}).

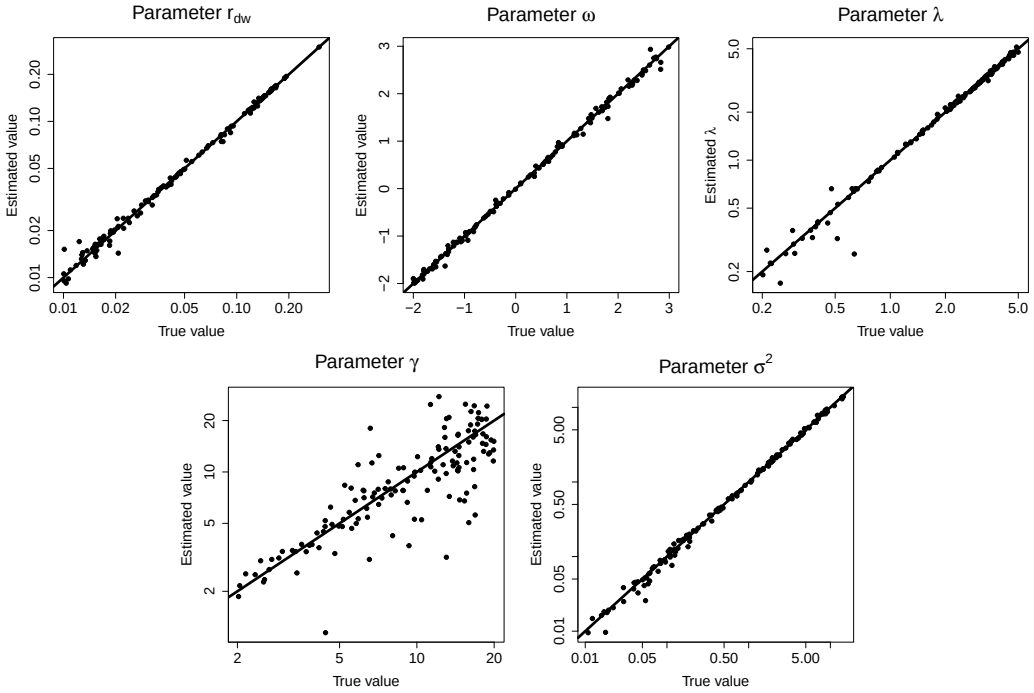


Figure S1: Practical identifiability for the dispersal model J_{Exp}. Each point represents the parameter estimation ('Estimated' value) depending on the real parameter ('True' value). Each graph regroups the results of 100 replicates. Straight lines correspond to the first bisector.

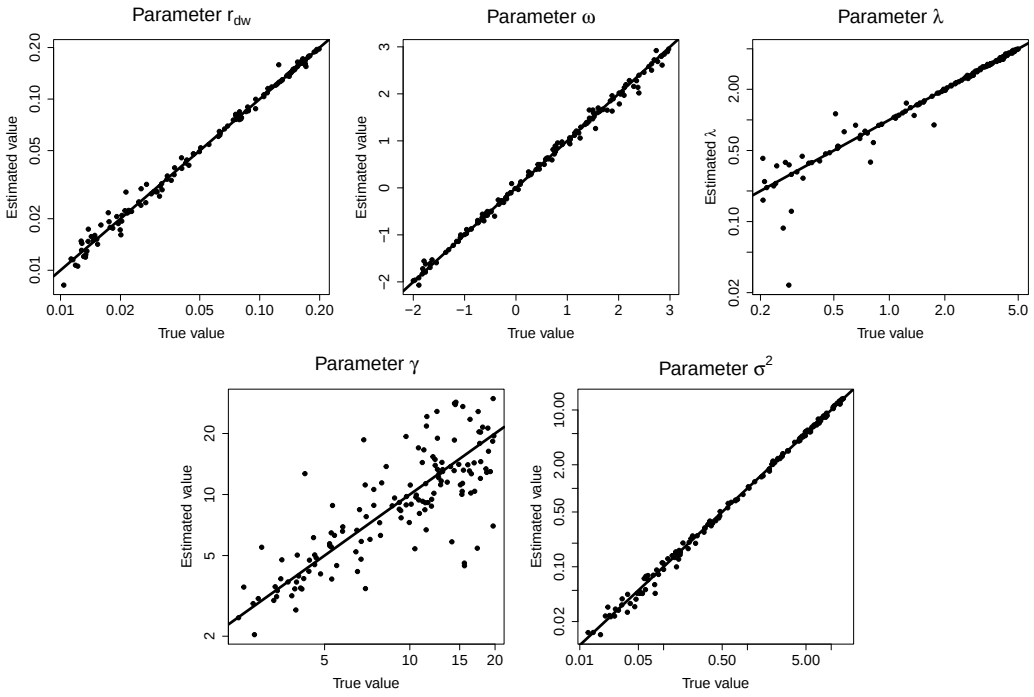


Figure S2: Practical identifiability for the dispersal model J_{Gauss}. Each point represents the parameter estimation ('Estimated' value) depending on the real parameter ('True' value). Each graph regroups the results of 100 replicates. Straight lines correspond to the first bisector.

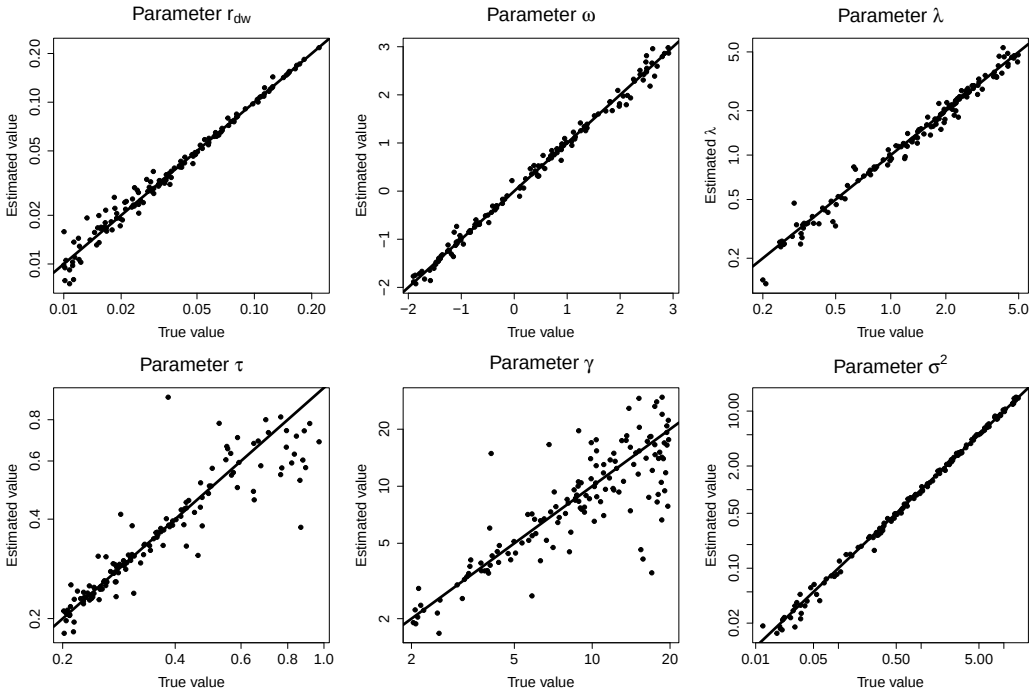


Figure S3: Practical identifiability for the dispersal model J_{ExpP}. Each point represents the parameter estimation ('Estimated' value) depending on the real parameter ('True' value). Each graph regroups the results of 100 replicates. Straight lines correspond to the first bisector.

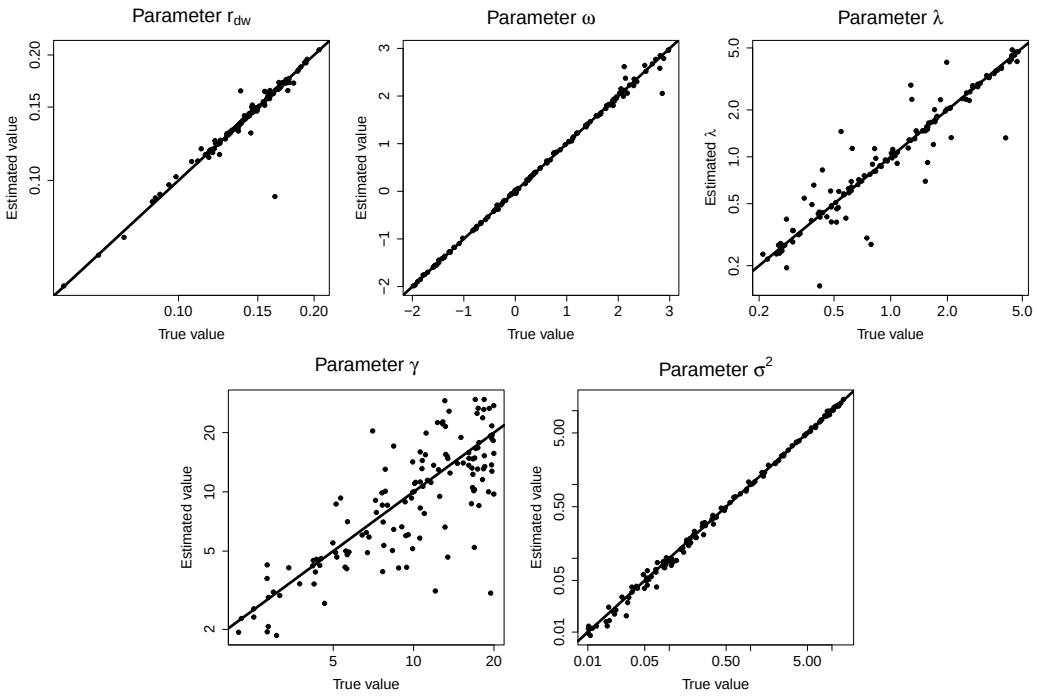


Figure S4: Practical identifiability for the dispersal model RD. Each point represents the parameter estimation ('Estimated' value) depending on the real parameter ('True' value). Each graph regroups the results of 100 replicates. Straight lines correspond to the first bisector.

D.2 Model selection

Model practical identifiability was similar to parameters practical identifiability (Appendix D.1), except that we fitted to each data set the true model (as previously) but also the 3 other models corresponding to the alternative hypotheses on dispersal process. Models were compared using AIC (Akaike Information Criteria) to select the best data-supported model. AIC were assessed as $2k - 2ln(L)$ where k is the number of parameters of the model considered and L is the maximized value of the likelihood function. To gain more insights into the confidence level in model selection, we also calculated for each data set the difference between the AIC of the model selected and the AIC of the second-best model according to the two possible issues of the selection procedure: (i) when the model selection procedure was successful (*i.e.* the selected model was the true model) and (ii) when the model selection procedure was incorrect (*i.e.* the true model was not selected). The mean of these values were reported as $dAIC_{true}$ and $dAIC_{wrong}$ in Table 2. The steps were reiterated until the estimation of $n = 50$ realistic epidemics for each dispersal model.

D.3 Parameter inference on the real data set

The model selection procedure was applied to the real data set by fitting four dispersal process hypotheses (J_{Exp}, J_{Gauss}, J_{ExpP} and RD). The same optimisation routines described in Appendix D.1 were performed from five initial conditions selected as in Step 3.2 (Appendix D.1). The selected model corresponds to hypothesis J_{ExpP}. We obtained maximum likelihood estimates of the vector of parameters $\hat{\theta}$ and of its matrix of variance-covariance $\hat{\sum}$. Confidence intervals were derived from 1,000 random draws from the multivariate normal distribution with parameters $\hat{\theta}$ and $\hat{\sum}$. The 95% confidence intervals of each parameter is obtained using the quantiles 2.5% and 97.5% (Table 4).

D.4 Model check

The model was checked by assessing the coverage rate of the data from the 95%-prediction intervals. The coverage rate was estimated as the proportion of observed data from the raw sampling within the prediction intervals (Figure 4).

Data from the raw sampling represent 97 counts Y_{st} of infected trees at sites $s \in \{1, \ldots, S\}$ (with $S = 12$ or $S = 45$ depending on the sampling date) and times $t \in \{1, \ldots, 6\}$. Let us recall that, as stated in Appendix B, Y_{st} follows a combination of Poisson and Beta-Binomial distributions whose parameters depend on the known mean value $(\lambda_m)_t$ and the unknown $u(t, x_s)$, γ and σ^2, and that $u(t, x_s)$ is a deterministic function of dynamical parameters r, λ and τ.

Prediction intervals were calculated at each date and each site with a two-step procedure:

Step 1 A confidence interval was obtained from 1000 random draw from the multivariate normal distribution with $\hat{\theta}$ and $\hat{\sum}$.

Step 2 The mean proportions of infected trees were calculated at each date and site date from each random draw of parameters obtained from Step 1. A prediction interval was obtained from these parameters given the probabilities of infection, with 1,000 random draws in the observation laws.

D.5 Model densification

As in Appendix D.2, numerical simulations were run to disentangle the true dispersal process from alternative dispersal processes, with densification of time and site for the raw and the refined sampling. Simulations were run with 21 sampling dates instead of 6, which amounts to one sampling every week. The number of sampling sites was set to 45 for all sampling dates. The steps described in Appendix D.1 and D.2 were reiterated until the estimation of $n = 50$ realistic epidemics for each dispersal model.

Table S3: Efficiency of model selection for the densification of time samples (21 instead of 6) and the site sampled (45 instead of 12). The four first columns indicate the proportion of cases, among 50 replicates, where each tested model was selected using AIC, given that data sets were generated under a particular model (*i.e.* true model). Column $dAIC_{true}$ (*resp.* $dAIC_{wrong}$) indicates the mean difference between the AIC of the model selected when the model selected is the true one (*resp.* when the model selected is not the true model) and the second best model (*resp.* being the true model or not).

	Selected Model					
True Model	J_{Exp}	J_{Gauss}	J_{ExpP}	RD	$dAIC_{true}$	$dAIC_{wrong}$
J_{Exp}	**0.72**	0.06	0.16	0.06	3.23	1.05
J_{Gauss}	0.22	**0.60**	0.04	0.14	7.33	1.67
J_{ExpP}	0.12	0.06	**0.82**	0	**1788.56**	2.72
RD	0.1	0.28	0.02	**0.60**	27.01	0.94

E Carrying capacity of poplar leaves

We measured the area of 10 wild poplar leaves (*Populus nigra*) and obtained a mean leaf area of $870\ mm^2$. We consider that poplar rust can not infect the leaf veins and edges, which represent approximately 15% of the leaf area. This leads to a net leaf area accessible to the pathogen of $740\ mm^2$. The size of a poplar rust lesion ranges from $0.2\ mm^2$ to $0.8\ mm^2$ (Maupetit et al., 2018). The lesions cannot fuse and are surrounded by living host tissue. We thus consider a lesion occupies a total area of $1\ mm^2$. This leads to a maximum of 740 lesions per leaf on average. To respect this order of magnitude, we consider in this analysis that the carrying capacity of a poplar leaf is 750 poplar rust lesions.

6.2 Dynamique d'expansion de *Melampsora larici-populina* dans la vallée de la Durance

À travers cette analyse, nous avons montré que notre méthode couplant un modèle mécanico-statistique à des données de comptage permet d'obtenir des estimations précises des paramètres régissant la dynamique d'expansion d'un agent pathogène.

En appliquant cette méthode aux données épidémiologiques de la colonisation de la vallée de la Durance par Mlp, nous obtenons la première estimation fiable de la dispersion chez cet agent pathogène. La distance moyenne de dispersion inférée par notre modèle est de 2,02 km, avec un intervalle de confiance de 95% allant de 1,76 à 2,27 km. Bien que peu d'estimations de ce type soient disponibles, l'ordre de grandeur de cette distance moyenne est cohérente avec les connaissances biologiques sur d'autres agents pathogènes fongiques. La distance moyenne de dispersion est estimée à 213 m pour les ascospores de *Mycosphaerella fijiensis*, 490 m pour les ascospores de *Leptosphaeria maculans*, 860 m pour *Podosphaera plantaginis* et 1380 à 2560 m pour *Hymenoscyphus fraxineus* (Fabre *et al.*, 2021b). Les estimations obtenues ici pour Mlp sont du même ordre de grandeur que celles obtenues pour *Hymenoscyphus fraxineus*, l'agent pathogène responsable de la maladie de la chalarose du frêne (Grosdidier *et al.*, 2018).

Au delà de la distance moyenne de dispersion, ce modèle nous permet de considérer différents régimes de dispersion qui peuvent être impliqués dans la colonisation. La fonction de dispersion inférée est celle donnée par le noyau Exponentiel-Puissance, avec un paramètre d'exposant du noyau de 0,24. Ce paramètre détermine la forme de la fonction de dispersion et cette faible valeur estimée indique des évènements de LDD particulièrement fréquents chez Mlp. Ces résultats sont conformes à ce que l'on peut attendre chez les Pucciniales, bien connues pour être dispersés par le vent sur de longues distances (Brown et Hovmøller, 2002; Aylor, 2003; Severns *et al.*, 2019).

Deux taux de croissance de Mlp ont été inférés de part et d'autre du barrage de Serre-Ponçon le long de la vallée de la Durance : 0.09 en amont et 0.02 en aval. Cette différence de taux de croissance correspond à un changement dans la topologie de la vallée. La vallée de la Durance est particulièrement encaissée et forme un corridor écologique étroit en amont, tandis qu'elle s'évase pour une ripisylve de peupliers plus large en aval. Le taux de croissance de Mlp peut dépendre de la température et de l'humidité (Somda et Pinon, 1981). L'aval de la vallée présente un climat bien plus chaud et sec que l'amont. Ces différences de taux de croissances peuvent donc également être expliquées par des variabilités climatiques entre ces deux régions. Une version antérieure de ce modèle permettait d'étudier l'influence de la température (températures moyennes, minimales et maximales journalières) sur le taux de croissance, *via* différentes fonction de liens. Nous avons utilisé pour cela les données de Météo-France obtenues dans le cadre de la convention « Re-

cherche » INRAE/Météo-France via l'unité AGROCLIM d'Avignon. Malheureusement cela n'a pas amélioré significativement la puissance des estimations et n'a donc pas été retenu dans le modèle final. Ces données sont issues de stations météorologiques réparties le long de la vallée de la Durance, qui ne coïncident pas nécessairement avec nos sites d'observations. Nous faisons l'hypothèse que le taux de croissance peut dépendre de variations climatiques très localisées que ne nous permettent pas de prendre en compte les données météorologiques utilisées. En effet, le microclimat local sur les sites d'observation au bord de la rivière est beaucoup plus humide que le climat enregistré dans les stations météorologiques situées près des agglomérations dans la vallée de la Durance. Nous avons identifié en particulier deux paramètres météorologiques dont la prise en compte dans le modèle pourrait permettre d'améliorer les estimations. La connaissance du point de rosée pour chaque site, c'est à dire la température en dessous de laquelle la rosée se dépose naturellement sur les feuilles de peuplier, nous permettrait de connaître les périodes favorables à la germination des spores et donc à l'infection par Mlp. Cette information est cependant très coûteuse à obtenir pour chacun des sites échantillonnés et nécessiterait des relevés spécifiques sur le terrain. La connaissance de la rose des vents tout au long de l'épidémie nous permettrait de calculer les degrés de connexion entre sites d'après le déplacement des masses d'air (LEYRONAS *et al.*, 2018). Ces informations seront bientôt disponibles grâce à l'outil TROPOLINK qui est en développement au sein de l'unité de recherche BIOSP, INRAE.

Enfin, nous avons établi dans ce modèle les lois d'observation de l'infection sur les feuilles, les rameaux et les arbres lors du suivi épidémiologique. Cela nous a permis d'estimer deux paramètres de ces lois d'observation. Le premier paramètre représente la façon dont les probabilités d'infection des feuilles, perçues par la personne en charge de l'échantillonnage, diffèrent d'un arbre à l'autre des probabilités réelles. Ce paramètre est estimé à 5,07, ce qui indique une variabilité de perception non négligeable mais modérée. Ces différences de perception de l'infection peuvent notamment être dues à la configuration spécifique du feuillage de chaque arbre. Le second paramètre représente la différence de réceptivité à la rouille d'une feuille à l'autre, due à l'hétérogénéité environnementale. Les feuilles observées le long de la ripisylve peuvent ainsi être plus ou moins humides ou présenter des variations locales d'exposition, notamment selon leur position sur l'arbre et l'éloignement à la rivière. Ces variations environnementales très localisées entraînent différents taux de germination des spores de rouilles à la surface des feuilles. Ce paramètre est estimé à 1,1, ce qui représente une forte variabilité de la réceptivité des feuilles à Mlp sur les sites échantillonnés.

La fréquence des évènements de LDD peut fortement impacter les structures génétiques sur un front de colonisation. La mise en évidence d'évènements de LDD particulièrement fréquents lors de la colonisation de la vallée de la Durance par Mlp nous amène

à nous demander quelles structures génétiques sont observées en fin de colonisation dans la vallée de la Durance et à quel point elles peuvent être expliquées par la dynamique d'expansion de l'agent pathogène.

Chapitre 7

Variabilité inter-annuelle de la structure génétique de *Melampsora larici-populina* dans la vallée de la Durance

L'objectif de ce chapitre est d'étudier la variabilité des structures génétiques observées empiriquement au cours de différents évènements de colonisation de Mlp dans la vallée de la Durance. Je souhaite notamment déterminer l'influence d'une dynamique d'expansion incluant des LDD, telle qu'inférée dans le chapitre précédent, sur les structures génétiques observées.

7.1 Article 6. Genetic signatures of a range expansion *in natura*: variation across years

Ce dernier chapitre a pour objet l'étude de la variabilité des structures génétiques sur un front de colonisation en comparant les isolats de Mlp prélevés en fin d'épidémie le long de la vallée de la Durance en 2004, 2008, 2009 et 2011. Nous étudions la variation de différents indices de génétique des populations le long de la vallée : 1) L'hétérozygotie théorique, 2) La richesse allélique (*i.e.* le nombre d'allèles pondéré par la taille de l'échantillonnage), 3) Le déséquilibre de liaison, 4) La richesse génotypique et 5) La différenciation génétique par rapport à la population source.

Ces résultats sont ici présentés sous la forme d'un article de recherche en préparation intitulé « Genetic signatures of a range expansion *in natura*: variation across years ».

Genetic signatures of a range expansion *in natura*: variation across years

Méline Saubin[1], Ronan Becheler[2], Axelle Andrieux[1], Pascal Frey[1], Solenn Stoeckel[2] and Fabien Halkett[1]

[1] Université de Lorraine, INRAE, IAM, F-54000 Nancy, France

[2] INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, F-35653 Le Rheu, France

1 Introduction

It has long been established that dispersal plays a dual role in shaping the genetic structures of populations. On the one hand, regular gene flow between populations via migrant individuals promotes genetic homogenisation of populations. On the other hand, colonists promote genetic differentiation by founding subpopulations that are genetically differentiated from the source population (Ibrahim et al., 1996, Austerlitz et al., 1997, Le Corre and Kremer, 1998). This duality can result in very different genetic structures on a colonisation front, depending on the relative strength of the dispersal mechanisms at play (Le Corre and Kremer, 1998; Fayard et al., 2009). In the case of a pushed wave, we expect conservation of neutral genetic diversity along the colonised domain, at least in the deterministic case (Bonnefon et al., 2014). This is because all individuals, regardless of their localisation, can contribute to the advance of the colonisation front, which maintains high effective population size throughout the colonisation wave (Fayard et al., 2009). Conversely, in the case of a pulled wave, rapid erosion of genetic diversity is expected as well as an increase in genetic differentiation from the source population. Indeed, only individuals at the forefront of the wave contribute to colonisation, hence maximising genetic drift and distorting the initial gene frequency (Edmonds et al., 2004; Wei and Krone, 2005). Pushed waves can lead to a steady increase in the frequency of an allele along the colonisation front, which can even lead to its fixation (Hallatschek et al., 2007). This phenomenon is called gene surfing, to depict the travel of such an allele along the expanding wave (Klopfstein et al., 2006; Excoffier and Ray, 2008). Many factors can moderate the resulting genetic structures between these two extremes (*e.g.* Allee effect, biotic interactions, climatic constraints, Miller et al., 2020). Among these factors, the dispersal of the species and the frequency of long distance dispersal events (LDD) may be decisive in explaining the genetic structures observed on a colonisation front (Ibrahim et al., 1996, Bialozyt et al., 2006, Fayard et al., 2009, Hallatschek and Fisher, 2014, Paulose and Hallatschek, 2020). Frequent LDD is expected theoretically to lead to conservation of genetic diversity (Bonnefon et al., 2014). This results from the regular arrival of new migrants on the colonisation front, thus preventing gene surfing. However, in

some cases long distance dispersal can lead to embolism effects (*e.g.* when dispersal occurs along a corridor, Fayard et al., 2009) or to patchy genetic structures, considerably reducing local genetic diversity (Paulose and Hallatschek, 2020).

Invasive species enable the joint study of both population demography and the resulting genetic structures on a colonisation front (Biek et al., 2007). Theoretical population genetics examines how the interplay between drift and dispersal regime shapes genetic diversity along range expansions. Yet it is challenging to test these predictions (such as the surfing phenomenon) in natural populations, in part because of the high stochasticity inherent to the realisation of each colonisation event. Here, we benefit from a remarkable ecological system to unravel the processes that rule colonisation patterns thanks to biological replicates. Embanked in the Alps, the Durance River valley (200km long and a few km wide) channels the dispersal of a plant pathogen, the poplar rust fungus (*Melampsora larici-populina*). Each year, this species spreads out clonally along the riverbanks. Waves of colonisation repeat on an annual basis starting from the same place of origin. Indeed, to complete its life cycle the poplar rust fungus has to infect two distinct host species: poplar (*Populus sp*, the telial host) and larch (*Larix sp*, the alternate host). This pathogen accomplishes a single sexual reproduction on larches in spring, then it infects poplars for a succession of clonal reproductions that are responsible for epidemics and expansion phases. The ecology and distribution of these two host species are well known (Riou Nivert et al., 2001; Imbert and Lefèvre, 2003; Pinon and Frey, 2005). Upstream the valley, larches and poplars coexist in a restricted area. In this area, the poplar rust pathogen can therefore easily switch hosts at the end of its sexual reproduction. This place forms the source population for recurrent epidemics in the Durance River valley (Xhaard et al., 2012). In summer, the poplar rust spreads downstream the valley as it colonises southward the young stands of wild poplars that are continuously distributed along the Durance River valley. The colonisation phase lasts until the fall of the poplar leaves that mark the end of the epidemic season in autumn. This invasion thus takes place every year from the same source location, in the same direction, along a one-dimensional and continuous domain. Each event of colonisation is a dead-end process since migrant individuals are too far from the poplar-larch sympatry area to be able to contribute to the sexual reproduction the following year.

A previous population genetics study by Xhaard et al. (2012) detected three genetically distinct groups of the poplar rust pathogen in the Durance River valley: the so-called 'wild', 'cultivated' and 'asexual' isolates. Only 'wild individuals' contribute to the epidemic initiated upstream of the Durance valley (Xhaard et al., 2012). The work of Xhaard et al. (2012) allows us to confidently assign a population of origin to each isolate. To investigate the neutral evolution of the poplar rust population through this recurrent colonisation event, we focus in this study on the wild type only. A study of the genetic structure resulting from the 2004 poplar rust invasion revealed a particular genetic structure, with neither conservation of genetic diversity nor allele surfing (Becheler et al., 2016). Instead, the clonal lineages that are repeatedly sampled along the river stands exhibit an erratic distribution, just as if they were playing 'leapfrog' from site to site, preventing a single genotype from surfing along the valley. A spatio-temporal epidemiological study of the 2008 invasion of the poplar rust pathogen inferred the dispersal of this pathogen as a fat-tailed dispersal kernel, *i.e.* with frequent LDD events (Article

5). This dispersal inference supports the findings of Becheler et al. (2016), with LDD events allowing clones to 'jump' over sites along the valley.

Here we aim at describing further the genetic structure resulting from LDD at the end of a colonisation event, by supplementing the study of Becheler et al., 2016 with 3 additional years of sampling. We aim to determine whether the observed genetic structures differ across years, and if so to what extend. In particular, we test for the conservation of initial gene diversity using a dedicated statistical method. We hence aim to provide, through the study of this remarkable ecological system, direct evidence of the interplay between drift and migration at work.

2 Materials and methods

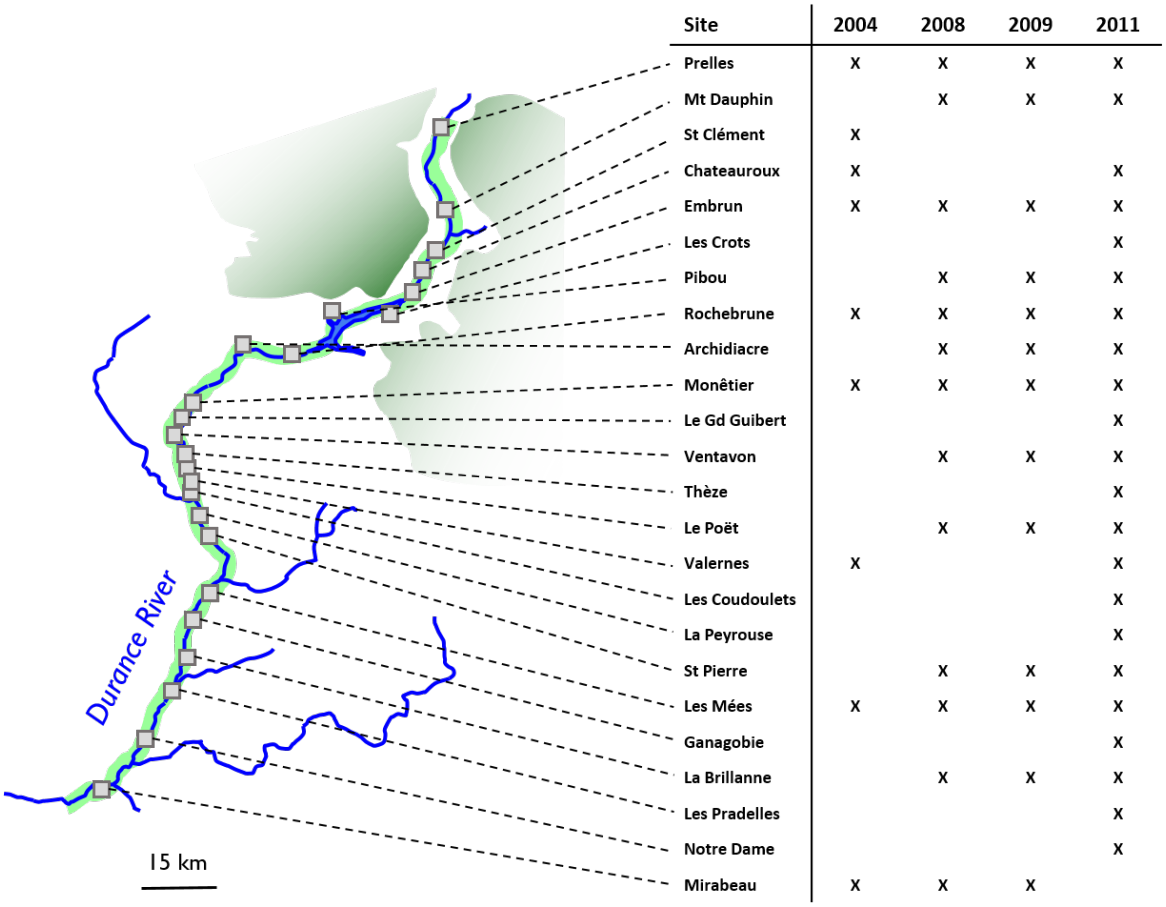
2.1 Sampling details

The study area is a corridor of 200 km long surrounded by wild poplars (*Populus nigra*) along the Durance River valley. A total of 24 sites are sampled for *Melampsora larici-populina* for four different years: 2004, 2008, 2009 and 2011 (Figure 1). Each sampled individual is isolated from a distinct leaf collected from different trees and twigs, to avoid resampling clonal replicates due to close proximity (Barrès et al., 2012). This study is a complete reanalyse of samples collected in 2004 (Xhaard et al., 2012; Becheler et al., 2016) genotyped with a new set of microsatellite markers, and supplements these results with the genotyping - using the same set of markers - of samples collected in 2008, 2009 and 2011.

Individuals collected a given year on a given site constitute a sampling. Four initial populations (one per year) are sampled at the beginning of the epidemic season, in Spring, on the most upstream site of the valley (Prelles). Fifty-seven populations are sampled on poplar leaves along the valley at the end of the epidemic season, in Autumn. This study analyses a total of 2,193 isolates from 61 populations: 10, 14, 14 and 23 populations from 2004, 2008, 2009 and 2011, respectively (Table 1). Thirteen populations are composed of less than 15 individuals because of low disease prevalence. To avoid biased interpretations due to low sample sizes, in most figures we plot these populations with shaded colours, and we do not plot populations composed of less than 5 individuals.

2.2 Microsatellite analysis

DNA of all individuals is extracted using a DNeasy96 DNA Plant kit (Qiagen) as described in Persoons et al., 2017. Microsatellite analyses are performed at 19 polymorphic loci: MLP12 (Barrès et al., 2008), MLP49, MLP50, MLP54, MLP55, MLP56, MLP57, MLP58, MLP66, MLP68, MLP71, MLP77, MLP82, MLP83, MLP87, MLP91, MLP93, MLP96, MLP97 (Xhaard et al., 2009, 2011). We use the same PCR protocol as described in Persoons et al. (2017).



Site	2004	2008	2009	2011
Prelles	X	X	X	X
Mt Dauphin		X	X	X
St Clément	X			
Chateauroux	X			X
Embrun	X	X	X	X
Les Crots				X
Pibou		X	X	X
Rochebrune	X	X	X	X
Archidiacre		X	X	X
Monêtier	X	X	X	X
Le Gd Guibert				X
Ventavon		X	X	X
Thèze				X
Le Poët		X	X	X
Valernes	X			X
Les Coudoulets				X
La Peyrouse				X
St Pierre		X	X	X
Les Mées	X	X	X	X
Ganagobie				X
La Brillanne		X	X	X
Les Pradelles				X
Notre Dame				X
Mirabeau	X	X	X	

Figure 1: Location of the 24 sites monitored on poplar leaves along the Durance River valley, with the table of sampling design for each year studied: 2004, 2008, 2009, and 2011. Dark green areas upstream represent the larch distribution; the pale green line along the valley represents riparian stands of wild poplars.

2.3 Population genetics analyses

2.3.1 Discrimination of clonal lineages

Melampsora larici-populina reproduces clonally on poplar leaves. The first step of the genotypic data analysis is to discriminate clonal lineages (multi-locus lineages, or MLLs). We identify and discriminate clonal lineages as described in Becheler et al. (2016). Based on these analyses, individuals identified with identical clonal lineages are confidently considered true clones.

2.3.2 Genetic assignation

Genotypes are assigned to the three genetic groups 'Asexual', 'Cultivated' and 'Wild' using ADEGENET (Jombart, 2008; Jombart and Ahmed, 2011). Individuals described in Xhaard et al. (2011) and assigned to these three genetic groups are used as references. The genotype partitioning using ADEGENET are designed to maximise the correspondence with previous assignments of reference individuals. We retain for this study only isolates assigned to the 'Wild' genetic group.

Table 1: Genetic characteristics of 57 populations of *Melampsora larici-populina* sampled on poplar leaves at the end of the epidemic season. N: number of individuals; R: clonal richness; MLL: number of distinct clonal lineages; A_r: allelic richness; H_E: expected heterozygosity; $\bar{r}_D$: multilocus index of linkage disequilibrium; F_{ST}: pairwise differentiation index between the given population and the initial population at Prelles the same year; $Fisher's\ Q-value$: Combined Fisher's exact test of conservation of locus-based genotype frequencies between the given population and the initial population at Prelles the same year.

Year	Site	N	R	MLL	A_r	H_E	$\bar{r}_D$	F_{ST}	$Fisher's\ Q-value$[a]
2004	Prelles	102	0.604	62	1.563	0.563	0.072	0	NA
2004	St Clément	10	0.778	8	1.539	0.539	0.162	0.018	0.094
2004	Chateauroux	13	0.083	2	1.474	0.474	1	0.12	< 0.0001 ***
2004	Embrun	97	0.573	56	1.567	0.567	0.076	0.006	< 0.0001 ***
2004	Rochebrune	44	0.651	29	1.559	0.559	0.071	0.014	< 0.0001 ***
2004	Monêtier	49	0.688	34	1.532	0.532	0.058	0.013	< 0.0001 ***
2004	Valernes	56	0.491	28	1.487	0.487	0.208	0.062	< 0.0001 ***
2004	Les Mées	19	0.556	11	1.495	0.495	0.248	0.064	< 0.0001 ***
2004	Mirabeau	2	1	2	1.377	0.377	NA	0.021	1
2008	Prelles	88	0.954	84	1.544	0.544	0.054	0	NA
2008	Mt Dauphin	45	1	45	1.547	0.547	0.021	0.001	0.994
2008	Embrun	44	1	44	1.553	0.553	0.001	-0.001	0.991
2008	Pibou	44	0.953	42	1.535	0.535	0.02	-0.001	0.957
2008	Rochebrune	42	0.902	38	1.559	0.559	0.077	0.005	0.539
2008	Archidiacre	39	1	39	1.563	0.563	0.049	0.007	0.712
2008	Monêtier	27	0.846	23	1.558	0.558	0.037	0.01	0.084
2008	Ventavon	26	0.76	20	1.562	0.562	0.099	0.01	0.054
2008	Le Poët	9	0.875	8	1.536	0.536	0.076	0.017	0.791
2008	St Pierre	30	0.69	21	1.52	0.52	0.113	0.016	0.069
2008	Les Mées	29	1	29	1.545	0.545	0.013	0.006	0.68
2008	La Brillanne	20	0.947	19	1.543	0.543	0.044	0.008	0.48
2008	Mirabeau	26	0.64	17	1.608	0.608	0.192	0.027	< 0.001 **
2009	Prelles	59	1	59	1.541	0.541	0.026	0	NA
2009	Mt Dauphin	49	1	49	1.56	0.56	-0.004	0.002	0.737
2009	Embrun	50	0.98	49	1.559	0.559	0.005	0	0.791
2009	Pibou	59	0.966	57	1.556	0.556	0.004	0.005	0.523
2009	Rochebrune	61	0.967	59	1.561	0.561	0.016	0.005	0.755
2009	Archidiacre	58	0.86	50	1.515	0.515	0.047	0.008	0.074
2009	Monêtier	59	0.724	43	1.532	0.532	0.076	0.011	< 0.001 **
2009	Ventavon	74	0.863	64	1.547	0.547	0.038	0.009	0.095
2009	Le Poët	82	0.901	74	1.553	0.553	0.054	0.005	0.452
2009	St Pierre	49	0.792	39	1.564	0.564	0.106	0.021	< 0.0001 ***
2009	Les Mées	38	0.865	33	1.526	0.526	0.032	0.014	0.075
2009	La Brillanne	82	0.84	69	1.548	0.548	0.032	0.013	0.002 *
2009	Mirabeau	62	0.443	28	1.53	0.53	0.256	0.047	< 0.0001 ***
2011	Prelles	36	0.971	35	1.544	0.544	0.013	0	NA
2011	Les Crots	42	0.78	33	1.547	0.547	0.098	0.006	0.943
2011	Mt Dauphin	30	0.966	29	1.557	0.557	0.039	0.002	1
2011	Chateauroux	7	1	7	1.537	0.537	0.001	-0.013	1
2011	Embrun	29	1	29	1.536	0.536	0.009	-0.001	1
2011	Pibou	30	1	30	1.539	0.539	0.007	0.005	1
2011	Rochebrune	30	0.862	26	1.56	0.56	0.088	-0.001	0.978
2011	Archidiacre	31	1	31	1.556	0.556	0.005	0.001	1
2011	Monêtier	28	0.926	26	1.55	0.55	0.034	0	1
2011	le Gd Guibert	35	0.882	31	1.553	0.553	0.028	-0.004	1
2011	Ventavon	27	0.962	26	1.541	0.541	0.021	0.016	0.964
2011	Thèze	30	0.759	23	1.52	0.52	0.137	0.018	0.698
2011	Le Poët	30	0.897	27	1.55	0.55	0.04	0.001	1
2011	Valernes	10	0.444	5	1.475	0.475	0.401	0.029	0.874
2011	Les Coudoulets	32	0.839	27	1.567	0.567	0.07	0.017	0.801
2011	La Peyrouse	18	0.706	13	1.611	0.611	0.139	0.014	0.855
2011	St Pierre	14	0.923	13	1.555	0.555	0.059	0.004	1
2011	Les Mées	2	1	2	1.456	0.456	NA	0.019	1
2011	Ganagobie	1	NA	1	1.444	0.444	NA	0.047	1
2011	La Brillanne	6	0.8	5	1.515	0.515	0.185	0.014	1
2011	Les Pradelles	5	1	5	1.513	0.513	-0.021	0.033	1
2011	Notre Dame	4	1	4	1.562	0.562	0.016	-0.001	1

[a] *P-values* for conservation of genotype frequencies are corrected for multiple tests and combined overall loci (*, $Q-value < 0.01$; **, $Q-value < 0.001$; ***, $Q-value < 0.0001$).

2.3.3 Population genetics indices

We analyse the population genetic structure with a discriminant analysis of principal components (DAPC, Jombart et al., 2010) as implemented in the R package ADEGENET. To illustrate genetic variations between upstream and downstream, we perform this analysis on populations at Prelles at the beginning and the end of the epidemic, and on populations downstream of Les Mées at the end of the epidemic for the four sampled years.

For each year and each location, we compute a set of population genetics indices using R (R, 2018). The clonal richness and linkage disequilibrium are computed with the R package POPPR (version 2.9.3, Kamvar et al., 2014, 2015). Clonal richness is estimated by $R = \frac{G-1}{N-1}$, with G the number of clonal lineages and N the number of individuals. Linkage disequilibrium between loci is estimated by $\bar{r}_D$ (Agapow and Burt, 2001). Gene diversity within locations and genetic differentiation among locations are computed with the R package HIERFSTAT (Goudet and Jombart, 2021). Gene diversity is estimated by the unbiased expected heterozygosity H_E (Nei, 1978). Genetic differentiation is estimated by the pairwise differentiation index F_{ST} between the site upstream (Prelles) at the end of the epidemic and each other site from the same year. For these last two indices, two series of genetic analyses are performed: 1) with all individuals; and 2) excluding clonal replicates by retaining a single copy of each clonal lineage in each site (as in Becheler et al., 2016).

2.3.4 Statistical analyses

We test for linkage disequilibrium in populations in Prelles at the beginning and the end of the epidemic, *i.e.* for populations sampled in spring and autumn. We exclude clonal replicates from each population by retaining a single copy of each clonal lineage and perform tests for linkage disequilibrium as implemented in the R package POPPR, with 999 permutations. We adjust the $P-values$ for multiple testing with the false discovery rate (FDR) procedure (Benjamini and Yekutieli, 2001). This procedure controls the proportion of significant results instead of controlling the chance of making even a single error. Adjusted $P-values$ are called $Q-values$. No population in Prelles present a significant linkage disequilibrium, at a significance threshold of 1% (Appendix, Table S1). This is consistent with the study of Xhaard et al. (2011). Therefore, at the start of the epidemic (*i.e.* after the sexual reproduction on the alternate host), all loci are initially independent.

To test for the hypothesis of pushed colonisation wave, we perform pairwise Fisher's exact tests on genotype frequencies at each locus between populations in Prelles and populations downstream, each year at the end of the epidemics. These tests are performed using the `fisher.test` R function with a two-sided alternative hypothesis and a significance threshold of 1%. We adjust the $P-values$ with the FDR procedure. We summarise the information on the distributions of $Q-value$ within each population using two metrics. First, we combine the resulting $Q-values$ with Fisher's method as implemented in the `fisher` R function to obtain a unique test result for each population. This amounts to strengthening the differences in genotypes' frequency as a single significant $Q-value$ is sufficient to drive combined $Q-values$ towards significance. As such, it provides a conservative test for the evenness of genotype frequency distribution over the colonised domain. Second, for each population,

we calculated the proportion of loci leading to a significant Fisher's exact test.

To test for the evolution of genotype frequencies at each locus over the distance, we implement multinomial logistic regressions and compute the corresponding $P-values$. We adjust the $P-values$ as above with the FDR procedure. We compare across years the proportion of loci leading to significant changes in genotypic frequencies with the distance, with a significance threshold of 1%. To explicitly test for the hypothesis of pulled colonisation wave, we perform linear regressions between the abundance of each repeated clonal lineage (when sampled at multiple locations) as a function of the distance to Prelles.

3 Results

3.1 Genetic evolution is independent across years

Prelles is the first site in the Durance valley to be infected, due to the coexistence of poplars and larches. None of the populations sampled at Prelles at the beginning and end of the epidemic display significant linkage disequilibrium between loci (Table S1). Therefore, all loci are initially independent at the beginning of the epidemic (*i.e.* after sexual reproduction on the alternate host), and this independence persists in Prelles until the end of the epidemic season. Populations at Prelles display similar distribution of genetic variation (Figure 2), highlighting the homogeneity of allelic frequencies across years. Conversely, downstream populations display differences in genetic structures depending on the year. Each annual epidemic hence represents an independent genetic evolution from a homogeneous inoculum source.

3.2 Variations in population genetics indices along the valley

Among all populations, we identify the population sampled at Chateauroux in 2004 as an outlier (Appendix, Figure S1). This population is composed of 13 individuals in two different clonal lineages only (Table 1), and is the only population composed exclusively of resampled clones. For clarification and to avoid graphical flattening of genetic variation at other sampling sites, we exclude this population from graphical representations of population genetics indices variations.

We observe a decrease in genotypic diversity along the valley each year, except in 2004 (Figure 3). However, in contrast to the other years, the initial genotypic diversity (at Prelles) in 2004 is already quite low. This erosion of genotypic diversity is associated with an increase in linkage disequilibrium along the valley, also visible in 2004. This illustrates the establishment of predominant clonal lineages along the colonisation front. Genetic differentiation also increases along the valley each year, but with differences in magnitude (Figure 3). This increase in genetic differentiation is less pronounced without replicates of clonal lineages.

Allelic richness and gene diversity are strongly correlated over all years (Appendix, Figure S1, correlation coefficient of 0.866). The year 2004 further stands out with a decrease in gene diversity and allelic richness along the valley (Figure 4). The other years do not show any particular trend in the change in gene diversity along the valley, except for 2008 with a high gene diversity at the last sampled

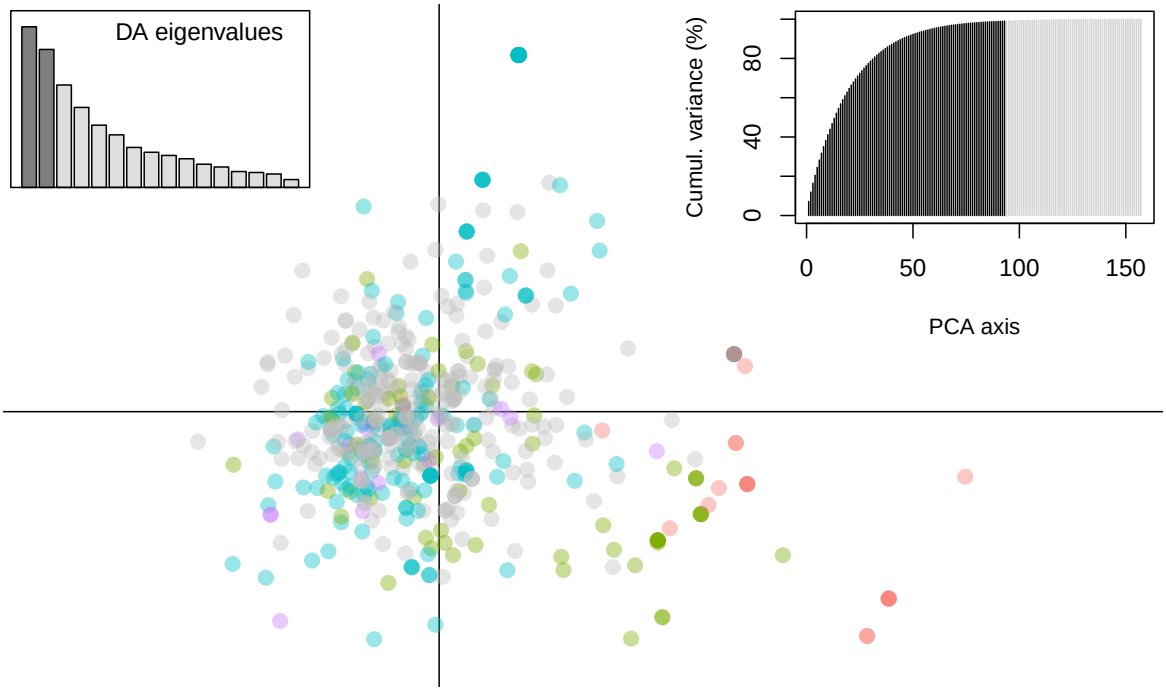


Figure 2: Scatterplot of the discriminant analysis of principal components (DAPC) for *M. larici-populina* genotypes collected along the Durance River valley at Prelles and downstream of Les Mées. The axes represent the first two linear discriminants. Grey dots represents isolates from Prelles collected in 2004, 2008, 2009 and 2011. Coloured dots represent isolates downstream of Les Mées collected in 2004 (red), 2008 (green), 2009 (blue), and 2011 (purple). DAPC analysis was performed with 93 first principal components (PCs) (99% of the variance conserved) of the principal component analysis (PCA) and 20 discriminant eigenvalues (DA)

site downstream. Similarly, allelic richness displays no clear trend along the valley, except for a slight decrease in 2009.

3.3 Evolution of genotype frequencies along the valley

To understand the genetic evolution of the *Melampsora larici-populina* along the colonisation front, we focus on two distinct scales: multilocus and locus-based evolution.

Focusing on the multilocus evolution, we observe distinct spatial patterns in the distribution of clonal lineages depending on the year. In 2004 and 2009 several clonal lineages were found in different and non-adjacent sites along the valley (Figure 5, three and two clonal lineages for 2004 and 2009, respectively). In 2004 we observe a particularly high proportion of clonal lineages resampled at the same sites. In 2008, one clonal lineage is sampled in two adjacent sites downstream the valley, while in 2011 no clonal lineages are sampled at different sites. This erosion of clonal diversity is consistent with the increase in population differentiation from Prelles accentuated in 2004 and 2009 (Figure 3). The frequencies of clonal lineages resampled at different locations are not correlated with the sample size, indicating no bias between the sampling effort and the clonal distribution (Appendix, Table S2). Among the six clonal lineages resampled inter-sites, five display no correlation between their frequency and the distance from Prelles (Table S2). However, for one clonal lineage sampled in 2004 at St Clément,

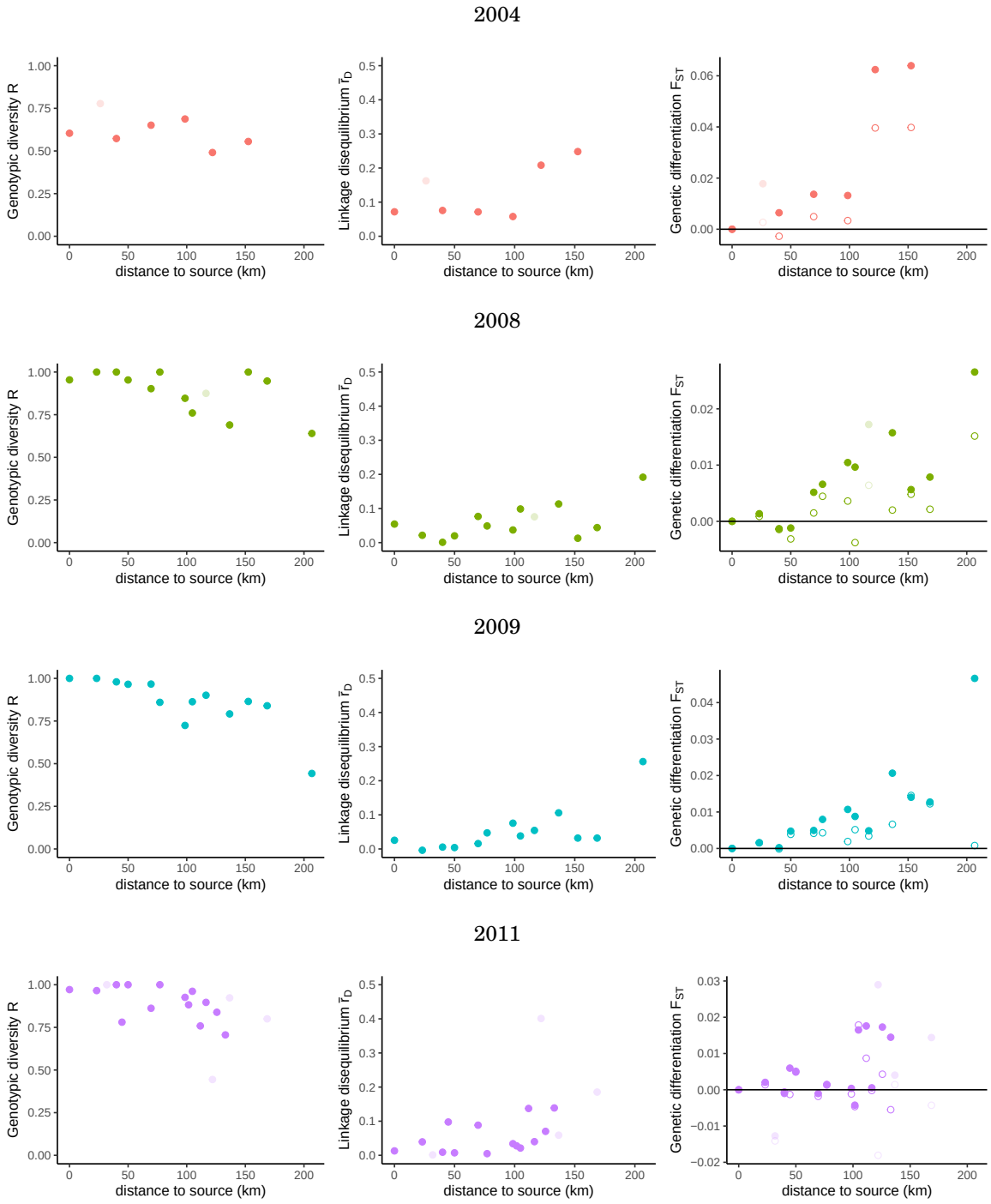


Figure 3: Evolution of the genotypic diversity (R), the linkage disequilibrium ($\bar{r}_D$), and the population differentiation from Prelles (pairwise F_{ST} with the population at Prelles as reference) with the distance from Prelles, for each monitored year. Dots represent indices calculated from complete populations; Circles represent indices calculated from populations restricted to one single replicate of each clonal lineage. Transparent dots represent populations of less than 15 individuals.

Monêtier, Valernes, Les Mées, and Mirabeau ($MLL1_{2004}$), there is a significant correlation between its frequency in the population and the distance from Prelles. This indicates that this clonal lineage could

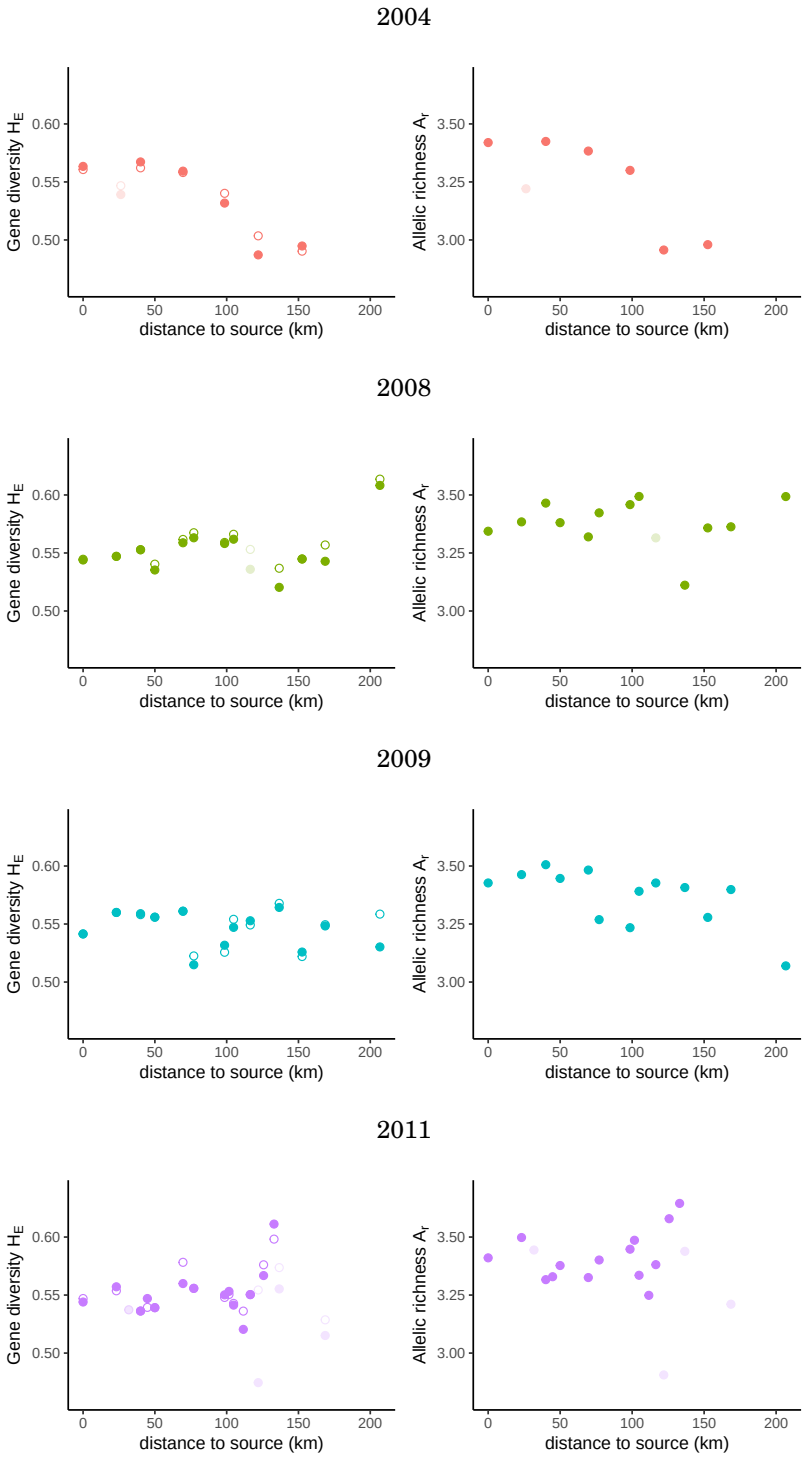


Figure 4: Evolution of the gene diversity (H_E) and the allelic richness (A_r) with the distance from Prelles, for each monitored year. Dots represent indices calculated from complete populations; Circles represent indices calculated from populations restricted to one single replicate of each clonal lineage. Transparent dots represent populations of less than 15 individuals.

have surfed on the colonisation front, as expected under the pulled-wave hypothesis. For this clonal

lineage, we test whether the lack of its detection between St Clément and Monêtier could be due to chance. Assuming an even distribution of this clonal lineage along the colonised domain, the probability to sample this lineage follows a Binomial law: $X \sim B(n, p)$, with X the occurrence of $MLL1_{2004}$, n the sample size, and p the frequency of $MLL1_{2004}$. We set the global frequency of $MLL1_{2004}$ at 0.038 (15 occurrences over 392 samples). The probability that we sample by chance no individual belonging to $MLL1_{2004}$ between Chateauroux and Rochebrune ($n = 154$ sampled individuals) is $P_B[X = 0] = 0.0026$. Thus, the lack of detection of $MLL1_{2004}$ between St Clément and Monêtier does not result from the sample size. While this clonal lineage can have surfed along the colonisation front downstream of Monêtier, this test provides evidence for at least one leapfrog dispersal event between St Clément and Monêtier.

Focusing on the locus-based evolution, the independence of allelic frequencies from the homogeneous inoculum source at Prelles enables to test for conservation in locus-based genotype frequencies along the colonised domain, as expected under the pushed-wave hypothesis. The results of Fisher's exact test enable to rank years between two extremes. In 2004, the locus-based genotype frequencies at Prelles are not conserved in any of the sampled sites, with a large proportion of loci displaying genotype frequencies significantly different from Prelles (Figure 6). In 2009, while the upstream sites show no significant differences from Prelles, four downstream sites differ, along with an increasing proportion of loci for which the genotype frequency is significantly different from Prelles. In 2004 and 2009, a large proportion of loci displays a significant correlation between their genotype frequencies and the distance from Prelles (Appendix, Table S3, 0.79 and 0.58 for 2004 and 2009, respectively). In 2008 and 2011, the locus-based genotype frequencies are conserved along the valley, except for the most downstream site in 2008. For both years, no locus shows a significant difference in genotype frequencies with Prelles (Figure 6), and the proportion of loci displaying a significant correlation between their genotype frequencies and the distance are low (Appendix, Table S3, 0.26 and 0 for 2008 and 2011, respectively). These results are consistent with the clonal lineages resampled along the valley in 2004 and 2009, resampled only downstream in 2008, and not resampled at all in 2011.

Thus, except for 2011, we do not necessarily observe the conservation of genotypic diversity of the inoculum source expected under the pushed-wave hypothesis. However, these changes in genotype frequencies along the colonised domain do not lead to fixation as expected under the pulled-wave hypothesis either.

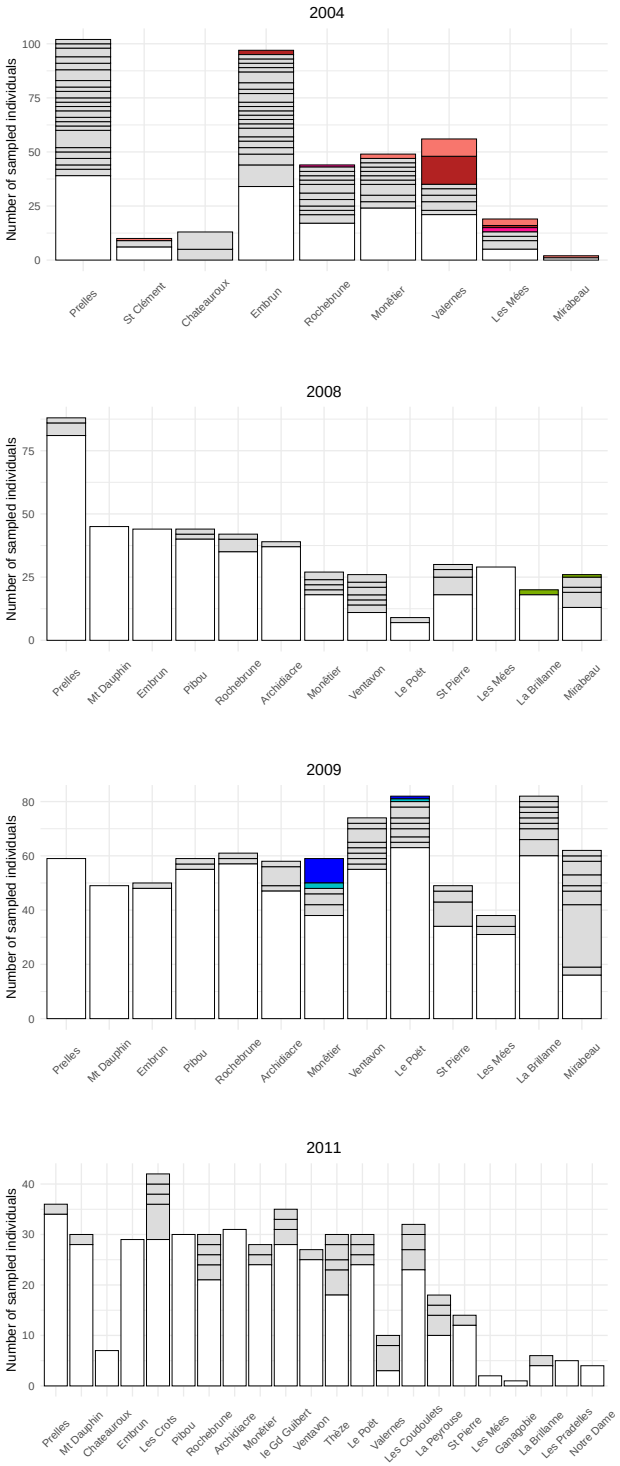


Figure 5: Details on the clonal composition of the sampled populations for each year and each location. In white the number of unique clonal lineages, (*i.e.* found in a single copy overall locations); In grey the number of locally abundant clonal lineages (*i.e.* multiple clonal copies found in a single site), each grey box represents the same clonal lineage; In colours the number of clonal lineages found in distinct sites, each colour represents a different clonal lineage.

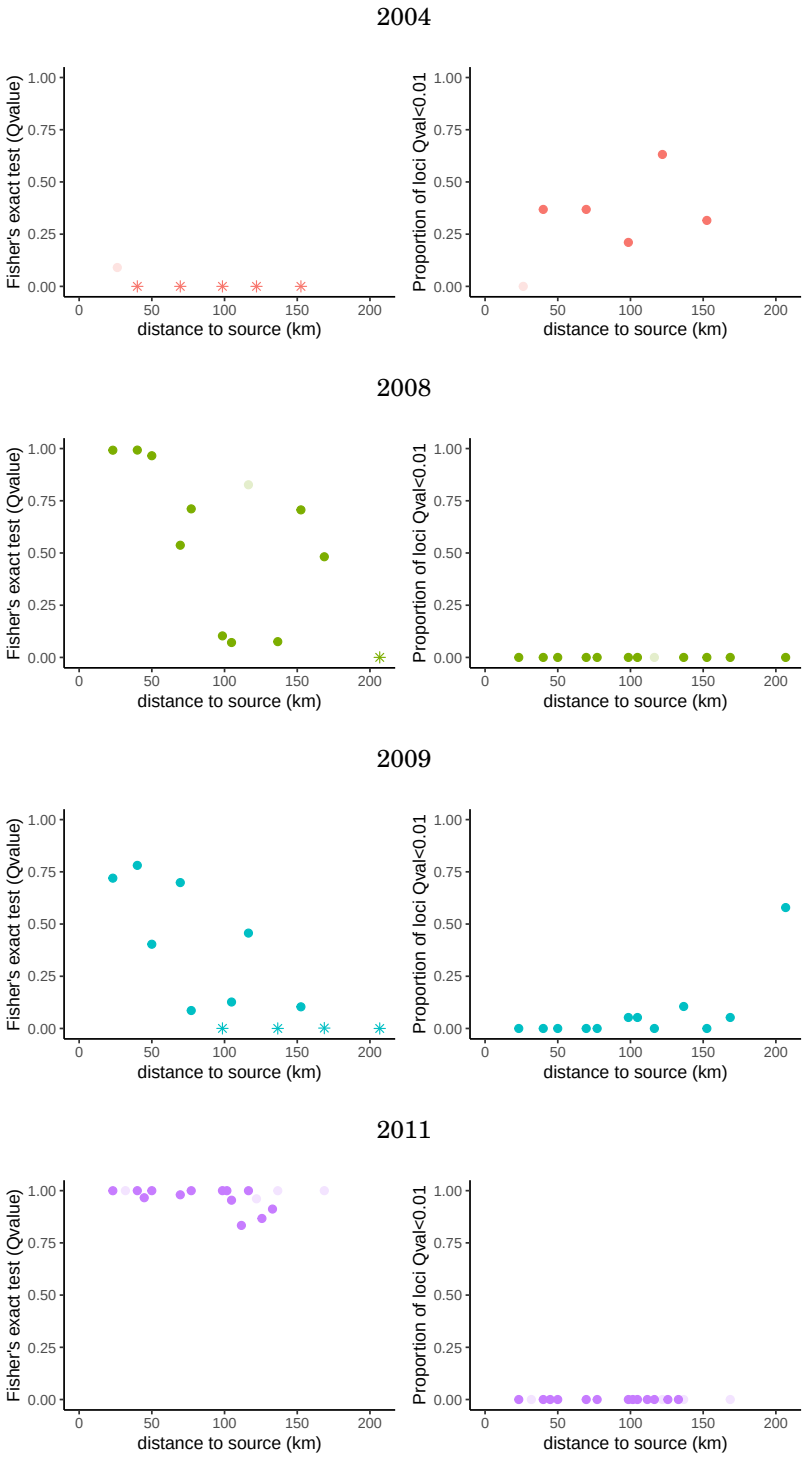


Figure 6: Fisher's exact test of conservation of genotype frequencies depending on the distance to Prelles. On the left, Fisher's exact test $Q-values$ combined over all loci: dots represent non significant $Q-values$, stars represent significant $Q-values$. On the right, proportion of loci with significant Fisher's exact test $Q-values$. The significance threshold is fixed at 1%. Transparent dots represent populations of less than 15 individuals.

4 Discussion

The typical genetic structure is shaped by LDD events

The colonisation of the Durance River valley by clonal epidemic waves of the poplar rust leads to a typical pattern of changes in population genetics indices along the colonised domain. This genetic signature of range expansion always involves a decrease in genotypic diversity that goes hand in hand with an increase in population differentiation and to a less extent an increase in linkage disequilibrium. Depending on the year considered, we also observe a decrease in gene diversity and allelic richness. All these signatures are the hallmark of recurrent founding events (Le Corre and Kremer, 1998) that result from the progressive colonisation of the valley (Xhaard et al., 2012, Article 5). Despite this regular change in population genetic structure along the colonised domain, we do not provide clear evidence for genotype surfing, as we find only one clonal lineage that displays a significant increase in frequency along the valley. Conversely, the other five clonal lineages to be resampled across sites are significantly non-evenly distributed and the other clonal lineages predominate only locally, which supports the finding of leaf frog events of dispersal (Becheler et al., 2016). In addition, for the year we do not observe leapfrog, there is a strong conservation of genotype frequencies along the colonised domain. This result is in accordance with previous estimates of the dispersal of rust fungi (Barrès et al., 2008; Brown and Hovmøller, 2002, Article 5), which suggests substantial long-distance dispersal (LDD) events. Notably, a previous epidemiological study provides detailed estimates of the dispersal capacity of *M. larici-populina* in the Durance River valley. This study is based on a spatio-temporal disease survey conducted in 2008 and uses a mechanistic-statistical approach to test for the dispersal regime that best explains the dynamics of colonisation. This work concludes that *M. larici-populina* disperses according to a fat-tailed Exponential-power kernel, with a mean exponent parameter of 0.24 and a mean dispersal distance of 2.02 km (Article 5). The annual colonisation of the Durance River valley during the clonal epidemic phase of *M. larici-populina* resulting from a fat-tailed kernel leaves a typical imprint along the colonised domain.

Inter-annual variation in genetic structures would result from differences across years in the frequency of LDD events

Beyond the annually conserved pattern of genetic structure resulting from colonisation, we highlight substantial variations in the magnitude of genetic changes across years. The range of variation lies between two extreme scenarios depicted in year 2004 and 2011, respectively. On the one hand, year 2004 displays the highest number of leapfrog events and the only case of genotype surfing downstream the valley, which translates into the strongest differentiation from the source population at Prelles. Year 2004 is also the only one to display a decrease in gene diversity and allelic richness along the colonised domain. On the other hand, year 2011 displays the weakest variation in population genetics indices, with no clonal lineages resampled across sites and no change in genotype frequency along the colonised domain. The genetic changes for years 2009 and 2008 rank between these two extremes. The four years differ in the magnitude of genetic changes but respect the same pattern described above, that is the differences across years affect all types of population genetics indices in the same way. These

differences in magnitude are perfectly illustrated by the test of conservation of genotypic frequencies which clearly ranks the years between the two extreme cases: from the conservation to the erosion of genotypic frequencies along the colonised domain.

A variation in the frequency of LDD events can explain the differences in the magnitude of genetic changes. If sufficiently frequent, LDD can contribute to the conservation of genetic diversity on a colonisation front and lead to a pushed wave (Bonnefon et al., 2014). Year 2011 is a typical example. Likewise, frequent LDD has been put forward to explain the lack of population genetic structure in a similar ecological system reporting the recent invasion of the plant pathogen *Pseudocercospora fijiensis* (Rieux et al., 2013). Conversely, if less frequent, LDD can no longer impede genotyping surfing nor the erosion of genetic diversity, especially in a corridor such as the Durance River valley, since a more gradual dispersal reduces the effective population size along the colonisation front (Fayard et al., 2009). In accordance, year 2004 displays a reduction of gene diversity along the colonised domain and is the sole year to provide evidence for a genotype surfing event. In this respect, our empirical results perfectly match the theoretical predictions of Bialozyt et al. (2006). In this simulation model, the authors show that the frequency of LDDs affects the distribution and the size of genotypic patches. LDD frequency thus determines the grain in a mosaic that figures genotypes' distribution. The more abundant the LDDs are, the finer the grain is, with many genotypes evenly distributed as in year 2011. The fewer the LDDs, the coarser the grain, with locally very abundant genotypes which can lead to the so-called embolism effect (Bialozyt et al., 2006; Petit, 2011). Differences in LDD frequency thus translate into differences in both genotype distribution and the magnitude of population genetics changes along range expansion. The more LDDs, the weakest the population genetic structure after range expansion.

Variations in ecological conditions lead to contrasted outcomes of the same colonisation process

We document here, for the same species and in the same ecological system, genetic structures along a colonisation front that are so different in magnitude that they range from the near absence of genetic changes to very large genetic differentiation caused by a founder effect by a handful of genotypes. Our empirical system well demonstrates the variability in the genetic structures that can be inherited from similar events of colonisation. As such it provides an illustration of the chance that comes with invasion events (Dlugosch and Parker, 2008; Desprez-Loustau et al., 2016), which challenges in return our ability to trace back invasion scenarios.

What causes such variations across years? We can provide at least two explanations that rely on differences in either dispersal or demography. First, as detailed above the patterns of population genetic structure are consistent with LDD events which frequency modulates the strength of the genetic signal. Inter-annual variation in population genetic structure can thus result from changes in dispersal regime caused by different climatic conditions (such as wind speed and direction). Second, the strength of the genetic signal is determined by the number of lineages that colonise the domain (Dlugosch and Parker, 2008). A limited number of founders causes a larger genetic differentiation (Le Corre and Kremer, 1998). Yet the peculiar life cycle of *M. larici-populina* can cause large differences in the initial

number of clonal lineages that spread down the Durance River valley, depending on how poplar rust switches hosts. Indeed the lack of synchrony between the bud burst on Larch and the emission of haploid spores from the fungal structure that overwinter on poplar leaves can cause a collapse in pathogen population size such as observed in other pathosystems, *e.g.* the oak mildew (Desprez-Loustau et al., 2010). The climatic conditions can play a key role in the degree of synchrony, with bud burst mainly determined by temperature and emission of haploid spores by air moisture. Such lack of synchrony may explain the typical signature of year 2004, with a reduced genotypic richness observed compared to the most upstream site of Prelles, where the host alternation takes place. Last, dispersal and demography can interplay. Genetic changes are linked to colonisation dynamics. A typical example, evidenced in (Becheler et al., 2016) is the site 'Chateauroux' in 2004, which was considered an outlier in the present study. Although it is located upstream of the valley this site displays a late colonisation from few clonal lineages. This results in a discrepant genetic structure with a much stronger genetic differentiation than expected for such an upstream location. This example illustrates the need to better couple population dynamics and genetics to further illuminate what causes population structure during colonisation.

5 Concluding remarks

Long-distance dispersal can leave a distinctive genetic imprint during colonisation, with an increase in genetic differentiation without gene surfing. We observe neither the conservation of genotypic diversity expected under the pushed-wave hypothesis nor genotype fixation expected under the pulled-wave hypothesis. Population genetic structures differ in magnitude depending on the interplay between dispersal regime and local demography. The observed genetic pattern should therefore allow to trace back the course of the colonisation and to test for the occurrence of LDD. Coupling these data with an appropriate demogenetic model would make it possible to build a sound and original framework to infer effective dispersal from genetic data. Collecting this information at fine spatial and temporal scales is crucial to propose relevant prophylactic or eradication policies to mitigate the impact of biological invasions and emerging diseases.

Acknowledgements

We warmly thank Frédéric Fabre and Jérôme Coville for insightful discussions on genetic evolution along a colonisation process. This work was supported by grants from the French National Research Agency (ANR-18-CE32-0001, CLONIX2D project; ANR-11-LABX-0002-01, Cluster of Excellence ARBRE). Méline Saubin was supported by a PhD fellowship from INRAE and the French National Research Agency (ANR-18-CE32-0001, CLONIX2D project).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agapow, P. M. and Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, 1(1-2):101–102.

Austerlitz, F., Jung-muller, B., Godelle, B., and Gouyon, P. H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, 51:148–164.

Barrès, B., Dutech, C., Andrieux, A., Halkett, F., and Frey, P. (2012). Exploring the role of asexual multiplication in poplar rust epidemics: Impact on diversity and genetic structure. *Molecular Ecology*, 21:4996–5008.

Barrès, B., Halkett, F., Dutech, C., Andrieux, A., Pinon, J., and Frey, P. (2008). Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, 8(5):577–587.

Becheler, R., Xhaard, C., Klein, E., Hayden, K. J., Frey, P., De Mita, S., and Halkett, F. (2016). Genetic signatures of a range expansion in natura: when clones play leapfrog. *Ecology and Evolution*, 6(18):6625–6632.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188.

Bialozyt, R., Ziegenhagen, B., and Petit, R. J. (2006). Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*, 19(1):12–20.

Biek, R., Henderson, J. C., Waller, L. A., Rupprecht, C. E., and Real, L. A. (2007). A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings of the National Academy of Sciences*, 104(19):7993–7998.

Bonnefon, O., Coville, J., Garnier, J., Hamel, F., and Roques, L. (2014). The spatio-temporal dynamics of neutral genetic diversity. *Ecological Complexity*, 20:282–292.

Brown, J. K. M. and Hovmøller, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science's compass*, 297:537–541.

Desprez-Loustau, M. L., Aguayo, J., Dutech, C., Hayden, K. J., Husson, C., Jakushkin, B., Marçais, B., Piou, D., Robin, C., and Vacher, C. (2016). An evolutionary ecology perspective to address forest pathology challenges of today and tomorrow. *Annals of Forest Science*, 73(1):45–67.

Desprez-Loustau, M. L., Vitasse, Y., Delzon, S., Capdevielle, X., Marçais, B., and Kremer, A. (2010). Are plant pathogen populations adapted for encounter with their host? A case study of phenological synchrony between oak and an obligate fungal parasite along an altitudinal gradient. *Journal of Evolutionary Biology*, 23(1):87–97.

Dlugosch, K. M. and Parker, I. M. (2008). Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, 17(1):431–449.

Edmonds, C. A., Lillie, A. S., and Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences*, 101(4):975–979.

Excoffier, L. and Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, 23(7):347–351.

Fayard, J., Klein, E., and Lefèvre, F. (2009). Long distance dispersal and the fate of a gene from the colonization front. *Journal of Evolutionary Biology*, 22(11):2171–2182.

Goudet, J. and Jombart, T. (2021). *hierfstat: Estimation and tests of hierarchical F-Statistics*.

Hallatschek, O. and Fisher, D. S. (2014). Acceleration of evolutionary spread by long-range dispersal. *Proceedings of the National Academy of Sciences*, 111(46):E4911–E4919.

Hallatschek, O., Hersen, P., Ramanathan, S., and Nelson, D. R. (2007). Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930.

Ibrahim, K. M., Nichols, R. A., and Hewitt, G. M. (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, 77:282–291.

Imbert, E. and Lefèvre, F. (2003). Dispersal and gene flow of *Populus nigra* (Salicaceae) along a dynamic river system. *Journal of Ecology*, 91(3):447–456.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24:1403–1405.

Jombart, T. and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*.

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11.

Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.*, 6:208.

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2:e281.

Klopfstein, S., Currat, M., and Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, 23(3):482–490.

Le Corre, V. and Kremer, A. (1998). Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology*, 11:495–512.

Miller, T. E. X., Angert, A. L., Brown, C. D., Lee-Yaw, J. A., Lewis, M., Lutscher, F., Marculis, N. G., Melbourne, B. A., Shaw, A. K., Szűcs, M., Tabares, O., Usui, T., Weiss-Lehman, C., and Williams, J. L. (2020). Eco-evolutionary dynamics of range expansion. *Ecology*, 101(10):1–14.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3):583–590.

Paulose, J. and Hallatschek, O. (2020). The impact of long-range dispersal on gene surfing. *Proceedings of the National Academy of Sciences*.

Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., and Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, 26(7):1902–1918.

Petit, R. J. (2011). Early insights into the genetic consequences of range expansions. *Heredity*, 106:203–204.

Pinon, J. and Frey, P. (2005). Interactions between poplar clone and Melampsora populations and their implications for breeding for durable resistance. *Rust Diseases of Willow and Poplar*, (July):138–154.

R (2018). R Core Team, R: A language and environment for statistical computing.

Rieux, A., De Lapeyre De Bellaire, L., Zapater, M. F., Ravigne, V., and Carlier, J. (2013). Recent range expansion and agricultural landscape heterogeneity have only minimal effect on the spatial genetic structure of the plant pathogenic fungus *Mycosphaerella* fijiensis. *Heredity*, 110(1):29–38.

Riou Nivert, P., Pâques, L. E., Le Pichon, C., and Bouvarel, L. (2001). *Le Mélèze*. Coll. Les guides du sylviculteur. IDF.

Wei, W. and Krone, S. M. (2005). Spatial invasion by a mutant pathogen. *Journal of Theoretical Biology*, 236(3):335–348.

Xhaard, C., Andrieux, A., Halkett, F., and Frey, P. (2009). Characterization of 41 microsatellite loci developed from the genome sequence of the poplar rust fungus, *Melampsora larici-populina*. *Conservation Genetics Resources*, 1(1):21–25.

Xhaard, C., Barrès, B., Andrieux, A., Bousset, L., Halkett, F., and Frey, P. (2012). Disentangling the genetic origins of a plant pathogen during disease spread using an original molecular epidemiology approach. *Molecular Ecology*, 21(10):2383–2398.

Xhaard, C., Fabre, B., Andrieux, A., Gladieux, P., Barrès, B., Frey, P., and Halkett, F. (2011). The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. *Molecular Ecology*, 20(13):2739–2755.

Appendix

Table S1: $Q-values$ resulting from the test for linkage disequilibrium at Prelles for populations sampled every year at the beginning (Spring) or the end (Autumn) of the epidemic. Tests were performed on populations without clonal replicates. MLL represents the number of distinct clonal lineages, *i.e.* the size of each population after clonal replicates were excluded.

Year	Time	MLL	$Q-value$
2004	Spring	4	0.286
2004	Autumn	62	0.286
2008	Spring	85	0.204
2008	Autumn	84	0.032 *
2009	Spring	68	0.868
2009	Autumn	59	0.803
2011	Spring	11	0.868
2011	Autumn	35	0.240

$P-values$ were corrected for multiple tests (*, $Q-value < 0.01$; **, $Q-value < 0.001$; ***, $Q-value < 0.0001$).

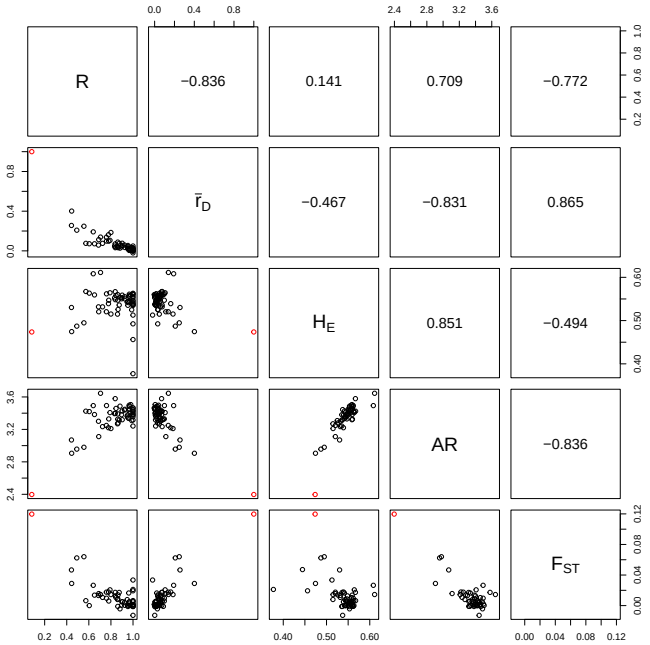
Table S2: Outcome of linear regressions between the proportions of each inter-site clonal lineage and the population size $(MLL \sim N)$, or between the proportions of each inter-site clonal lineage and the distance from Prelles $(MLL \sim distance)$.

Year	inter-site MLL ID	$(MLL \sim N)$		$(MLL \sim distance)$	
		$P-value$	r^2	$P-value$	r^2
2004	$MLL1_{2004}$	0.128	0.30	0.005 **	0.70
2004	$MLL2_{2004}$	0.759	0.014	0.457	0.081
2004	$MLL3_{2004}$	0.516	0.063	0.327	0.14
2008	$MLL1_{2008}$	0.313	0.092	0.050	0.31
2009	$MLL1_{2009}$	0.623	0.023	0.872	0.002
2009	$MLL2_{2009}$	0.964	0.0002	0.941	0.0005

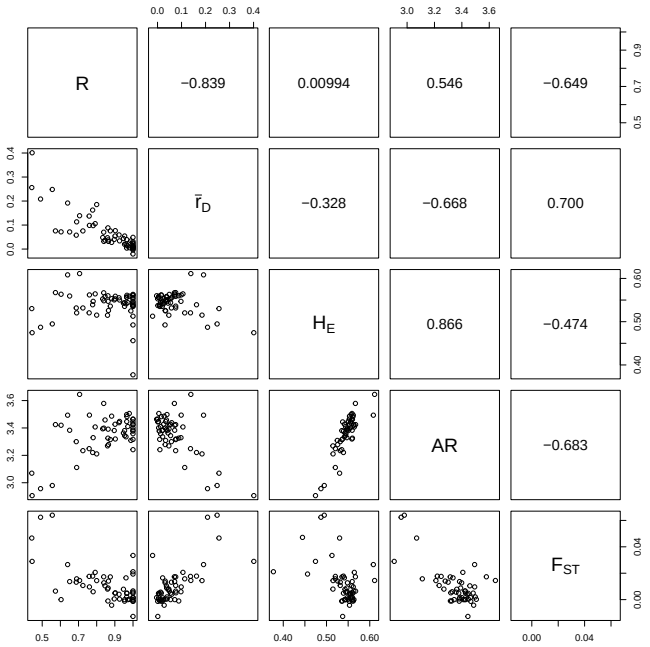
*, $P-value < 0.01$; **, $P-value < 0.001$; ***, $P-value < 0.0001$

Table S3: Outputs from the multinomial logistic regression to test for the dependency between locus-based genotype frequencies and distance from Prelles. The proportion of loci $Qval > 0.01$ represents the proportion of loci each year (among 19 studied loci) for with the regression was significant, with a significance threshold of 1%.

Year	Proportion of loci $Qval > 0.01$
2004	0.79
2008	0.26
2009	0.58
2011	0



(a) All populations



(b) Exclusion of the population sampled at Chateauroux in 2004

Figure S1: Co-variations between population genetics indices among sampled populations, (a) with or (b) without the outlier population sampled at Chateauroux in 2004. The upper matrix represent the correlation coefficients between indices. Each dot represent a population, the outlier population is represented by a red dot.

7.2 Dynamique d'expansion et structures génétiques

Cette analyse nous a permis de caractériser et comparer les structures génétiques de plusieurs évènements annuels de colonisation. Nous avons montré que les structures génétiques de Mlp en fin de colonisation sont similaires mais varient en amplitude d'une année sur l'autre. Ces variations en amplitude se traduisent en particulier par deux structures génétiques extrêmes : en 2011, nous observons une forte conservation de la diversité génétique initiale le long du domaine colonisé ; en 2004, nous observons à l'inverse une érosion rapide de la diversité génétique. Les années 2008 et 2009 présentent des structures génétiques intermédiaires.

Nous savons d'après le chapitre précédant qu'en 2008 au moins, les évènements de LDD chez Mlp étaient très fréquents lors de la colonisation de la vallée de la Durance. Ces évènements de LDD expliquent notamment la prédominance locale de certaines lignées clonales et l'existence d'autres lignées ré-échantillonnées sur plusieurs sites non consécutifs de la vallée. Cette structure forme une mosaïque de taches clonales caractéristiques d'évènements de LDD le long d'un corridor écologique (Bialozyt *et al.*, 2006).

Les variations inter-annuelles dans l'amplitude de ces changements génétiques peuvent être expliquées par des variations dans la fréquence de ces évènements de LDD. Pour des fréquences particulièrement fortes, les flux de gènes très importants entre l'avant et l'arrière du front ont pu mener à une forte conservation de la structure génétique telle qu'elle est observée en 2011. Pour une fréquence plus faible d'évènements de LDD, les flux de gènes ont pu être insuffisants pour compenser les effets de fondation répétés à l'avant du front de colonisation et mener à une structure génétique telle qu'observée en 2004.

Une autre campagne d'échantillonnage a été réalisée dans le cadre de cette thèse, à l'automne 2020. Les isolats ont été génotypés mais l'analyse n'est pas finalisée. Les résultats obtenus seront rajoutés à cette étude par la suite. D'après l'analyse préliminaire de ces génotypes, l'épidémie de l'année 2020 semble présenter une structure génétique proche de celle observée en 2009. Les isolats présentent de nombreuses lignées clonales prédominantes localement et au moins deux lignées clonales ré-échantillonnées dans plusieurs sites non consécutifs.

Conclusion de la partie II

La seconde partie de cette thèse a permis de mettre en évidence le lien entre la dynamique d'expansion et l'évolution génétique d'une population d'agents pathogènes responsables d'une maladie émergente. J'ai ainsi pu inférer les paramètres régissant la dynamique d'expansion de Mlp dans la vallée de la Durance et notamment mis en évidence l'existence de fréquents évènements de dispersion à longue distance. J'ai ensuite étudié les structures génétiques résultant d'une telle dynamique d'expansion et comparé les profils obtenus après plusieurs évènements de colonisation. J'ai montré que les structures génétiques observées sont caractéristiques d'évènements de dispersion à longue distance et que les différences dans leur amplitude peuvent être expliquées par différentes fréquences de ces évènements. Ce corridor écologique, où la composante spatiale est déterminante, a ainsi permis d'étudier l'évolution de populations d'agents pathogènes lorsqu'elles présentent de fortes variations démographiques.

Pour les années 2004, 2009 et 2011, les données épidémiologiques similaires à celles de la colonisation de 2008 ne sont pas disponibles. De telles données permettraient d'inférer les paramètres de la dynamique d'expansion chaque année. Grâce aux valeurs inférées, nous pourrions alors confronter ces inférences à l'hypothèse posée de la variation interannuelle de la fréquence des évènements de dispersion à longue distance. En l'absence de telles données cependant, une autre manière de vérifier cette hypothèse serait de compléter le modèle mécanico-statistique utilisé dans le Chapitre 3 en y ajoutant l'évolution de marqueurs génétiques neutres lors de la colonisation. Nous pourrions ainsi appliquer le même cadre d'analyse que celui utilisé dans le Chapitre 3 : 1) Construire un plan de simulation par la variation continue des paramètres de la dynamique d'expansion ; 2) Identifier les structures génétiques qui en résultent le long du domaine colonisé ; 3) Interpréter ces structures en regard des variations démographiques. Un tel modèle démogénétique permettrait de comprendre précisément l'influence de la dynamique d'expansion sur les structures génétiques observées. Faute de temps, ce suivi génétique n'a pas été implémenté dans le cadre de cette thèse mais pourra faire l'objet de futurs travaux de recherche.

Au delà de cette compréhension théorique, nous pourrions ensuite appliquer le même cadre d'analyse que celui utilisé dans le Chapitre 4 pour inférer les paramètres d'expansion à partir des données génétiques par une approche ABC. Il pourrait être intéressant de déterminer à quel point la fréquence et la régularité des échantillonnages dans le temps et dans l'espace impactent la qualité de l'inférence. Notamment, nous nous demandons si l'échantillonnage d'isolats seulement en fin d'épidémie peut apporter suffisamment d'informations génétiques pour retracer la dynamique d'expansion. Enfin, l'évolution génétique de la population peut permettre de distinguer plus finement les processus de démographie et de dispersion lors de chaque évènement de colonisation. Les éventuels écarts entre les données empiriques et le modèle théorique pourront ainsi être interprétés comme l'existence de mécanismes non pris en compte dans le modèle, tels que la sélection. Ces mécanismes pourraient alors être implémentés dans le modèle pour rendre compte plus précisément du processus biologique observé.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

La démogénétique établit un lien entre la génétique des populations et l'épidémiologie

Les disciplines de la génétique des populations et de l'épidémiologie sont historiquement peu liées, de par les échelles de temps et d'espace qu'elles considèrent, les questions auxquelles elles s'intéressent et les outils qu'elles utilisent pour y répondre (MILGROOM ET PEEVER, 2003). L'un des principaux objectifs de l'épidémiologie est d'établir une compréhension théorique des épidémies dans le temps et dans l'espace. L'objectif principal de la génétique des populations est de comprendre les processus évolutifs qui façonnent l'évolution génétique intra et inter populations. Je montre dans cette thèse que la démogénétique permet de prendre en compte à la fois les variations démographiques au centre des préoccupations en épidémiologie et les variations génétiques au centre de la génétique des populations. Cette prise en compte conjointe est possible car nous travaillons à la même échelle de temps et d'espace pour l'étude des processus démographiques et génétiques. Les outils démogénétiques établissent le lien entre ces disciplines pour réaliser des études de biologie des populations (MILGROOM, 2001; MILGROOM ET PEEVER, 2003) (FIGURE 7.1). La démogénétique permet ainsi de prendre en compte les rétroactions entre les processus écologiques et évolutifs, souvent considérés indépendamment les uns des autres (MILLER *et al.*, 2020), pour une meilleure compréhension des systèmes biologiques.

Dans chaque partie de cette thèse, j'étudie un système écologique pour lequel les variations démographiques et génétiques d'une population sont étroitement liées. Pour de tels systèmes, l'étude de ces variations indépendamment l'une de l'autre n'aurait pas permis de comprendre et encore moins de prédire le fonctionnement biologique de la population étudiée. Je montre notamment que ce ne sont pas les évènements démographiques les plus forts qui mènent nécessairement aux structures génétiques les plus marquées. Il peut exister des contournements de résistance sans signatures génétiques fortes (les trajectoires évolutives présentées dans le Cluster 1, Chapitre 3) et nous avons observé une invasion biologique sans différenciation génétique (Colonisation de la vallée de la Durance par Mlp en 2011, Chapitre 7). Cette thèse permet de mettre en évidence l'établissement très rapide de structures génétiques conjointement à des variations démographiques. Notre compréhension du lien entre les variations démographiques et l'évolution génétique reste cependant incomplète. Par la suite, il pourrait être intéressant de regarder plus précisément s'il existe une corrélation entre la vitesse d'expansion démographique, la vitesse d'établissement de ces structures génétiques et la magnitude de ces changements.

La démogénétique permet d'étudier des phénomènes contemporains

Les cadres d'analyses démogénétiques développés dans cette thèse ont permis de modéliser et d'étudier des variations démographiques et génétiques d'évènements contemporains. Ces études se distinguent des modèles classiquement utilisés pour retracer l'histoire

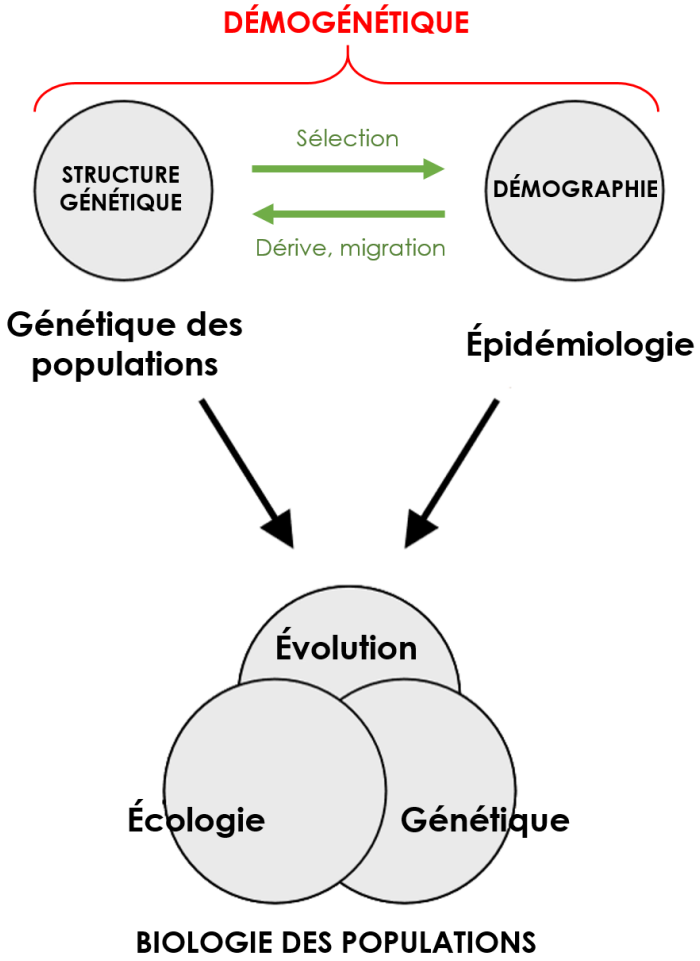


FIGURE 7.1 – Représentation schématique du lien entre la génétique des populations et l'épidémiologie pour former la biologie des populations, grâce aux outils de démogénétique (adapté d'après MILGROOM, 2001).

évolutive à long terme (BAZIN *et al.*, 2010; ROSENBERG ET NORDBORG, 2002; CORNUET *et al.*, 2010; ESTOUP ET GUILLEMAUD, 2010). De tels modèles contemporains et contextualisés sont particulièrement intéressants pour étudier des populations lors d'évènements d'adaptation rapide. Les agents pathogènes représentent pour cela de très bon modèles biologiques, étant sujets à une évolution rapide et contemporaine liée à de fortes variations démographiques (MCDONALD, 2010; PERSOONS *et al.*, 2017). Les populations d'agents pathogènes peuvent cependant être difficiles à étudier du fait de cette adaptation rapide et de la stochasticité des trajectoires (PARSONS *et al.*, 2018). L'approche démogénétique a fait ses preuves pour étudier de telles populations, à condition de disposer d'échantillons temporels suffisamment sériés pour caractériser les trajectoires évolutives des agents pathogènes, ce qui plaide en faveur de suivis de terrain sur le long terme. Cette approche peut donc représenter un coût considérable, notamment par l'effort d'échantillonage que cela implique. Le développement de réseaux d'épidémio-surveillance de plus en plus performants rend ce-

pendant de telles études plus accessibles. La surveillance régulière permet ainsi d'obtenir des échantillonnages temporels et d'identifier *a posteriori* les populations sources des épidémies. Bien que ces outils de terrain soit plus performants pour suivre les agents pathogènes en temps réel, ils restent majoritairement appliqués à un suivi épidémiologique des populations. Coupler ces suivis épidémiologiques à une caractérisation génétique des populations permettrait de mieux retracer les trajectoires évolutives des agents pathogènes. De tels suivis peuvent être réalisés par exemple par des prélèvements aérobiologiques réguliers couplés à des analyses moléculaires (AGUAYO *et al.*, 2021), permettant de détecter, quantifier et caractériser génétiquement les spores d'agents pathogènes fongiques dispersées par le vent à de larges échelles.

La démogénétique représente un outil pour répondre aux demandes sociétales actuelles en santé des plantes et des forêts

Enfin, ce cadre d'analyse démogénétique permet de comprendre les conséquences épidémiologiques des pratiques agricoles. C'est le cas, par exemple, de la mise en évidence dans le Chapitre 2 de l'impact de la proportion d'hôtes résistants déployés dans le paysage sur la probabilité et la vitesse d'un contournement de résistance. Pour aller plus loin dans la lutte contre les agents pathogènes, la démogénétique nous donne la possibilité de dépasser le cadre exploratoire et descriptif classique des études en écologie et en biologie de l'évolution (BERNER ET AMRHEIN, 2022), pour aller jusqu'à l'inférence des paramètres et la prédiction des trajectoires évolutives et démographiques des populations. De tels modèles contextualisés et prenant en compte les variations démogénétiques permettent ainsi d'étudier des populations contemporaines et de collecter une large gamme d'informations. Ces atouts sont déterminants pour renforcer les stratégies de lutte contre les agents pathogènes responsables de maladies émergentes.

Bibliographie

AGRIOS, G. N. (2005). *Plant pathology*. Elsevier édition.

AGUAYO, J., HUSSON, C., CHANCEREL, E., FABREGUETTES, O., CHANDELIER, A., FOURRIER-JEANDEL, C., DUPUY, N., DUTECH, C., IOOS, R., ROBIN, C., THIBAUDON, M., MARÇAIS, B. et DESPREZ-LOUSTAU, M. L. (2021). Combining permanent aerobiological networks and molecular analyses for large-scale surveillance of forest fungal pathogens : A proof-of-concept. *Plant Pathology*, 70(1):181–194.

ANDERSON, P. K., CUNNINGHAM, A. A., PATEL, N. G., MORALES, F. J., EPSTEIN, P. R. et DASZAK, P. (2004). Emerging infectious diseases of plants : pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution*, 19(10):536–544.

ARNAUD-HAOND, S., DUARTE, C. M., ALBERTO, F. et SERRÃO, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, 16:5115–5139.

AUSTERLITZ, F., JUNG-MULLER, B., GODELLE, B. et GOUYON, P. H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, 51:148–164.

AYLOR, D. E. (2003). Spread of plant disease on a continental scale : Role of aerial dispersal of pathogens. *Ecology*, 84(8):1989–1997.

BARRÈS, B., HALKETT, F., DUTECH, C., ANDRIEUX, A., PINON, J. et FREY, P. (2008). Genetic structure of the poplar rust fungus *Melampsora larici-populina* : Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, 8(5):577–587.

BAZIN, E., DAWSON, K. J. et BEAUMONT, M. A. (2010). Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics*, 185(2):587–602.

BECHELER, R., XHAARD, C., KLEIN, E., HAYDEN, K. J., FREY, P., DE MITA, S. et HALKETT, F. (2016). Genetic signatures of a range expansion in natura : when clones play leapfrog. *Ecology and Evolution*, 6(18):6625–6632.

BERNER, D. et AMRHEIN, V. (2022). Why and how we should join the shift from significance testing to estimation. *Journal of Evolutionary Biology*, 35(6):777–787.

BIALOZYT, R., ZIEGENHAGEN, B. et PETIT, R. J. (2006). Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*, 19(1):12–20.

BONNEFON, O., COVILLE, J., GARNIER, J., HAMEL, F. et ROQUES, L. (2014). The spatio-temporal dynamics of neutral genetic diversity. *Ecological Complexity*, 20:282–292.

BOUIN, E., GARNIER, J., HENDERSON, C. et PATOUT, F. (2018). Thin front limit of an integro-differential Fisher-KPP equation with fat-tailed kernels. *SIAM Journal on Mathematical Analysis*, 50(3):3365–3394.

BOUSSET, L., SPRAGUE, S. J., THRALL, P. H. et BARRETT, L. G. (2018). Spatio-temporal connectivity and host resistance influence evolutionary and epidemiological dynamics of the canola pathogen *Leptosphaeria maculans*. *Evolutionary Applications*, 11:1354–1370.

BRESSON, A., JORGE, V., DOWKIW, A., GUERIN, V., BOURGAIT, I., TUSKAN, G. A., SCHMUTZ, J., CHALHOUB, B., BASTIEN, C. et FAIVRE RAMPANT, P. (2011). Qualitative and quantitative resistances to leaf rust finely

mapped within two nucleotide-binding site leucine-rich repeat (NBS-LRR)-rich genomic regions of chromosome 19 in poplar. *New Phytologist*, 192(1):151–163.

BROWN, J. K. M. (2015). Durable resistance of crops to disease : a darwinian perspective. *Annual Review of Phytopathology*, 53:513–539.

BROWN, J. K. M. et HOVMØLLER, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science's compass*, 297:537–541.

BROWN, J. K. M. et TELLIER, A. (2011). Plant-parasite coevolution : Bridging the gap between genetics and ecology. *Annual Review of Phytopathology*, 49(1):345–367.

BROWNING, J. A. et FREY, J. K. (1969). Multiline cultivars as a means of disease control. 7:355–382.

BURDON, J. J., BARRETT, L. G., REBETZKE, G. et THRALL, P. H. (2014). Guiding deployment of resistance in cereals using evolutionary principles. *Evolutionary Applications*, 7(6):609–624.

BURDON, J. J., ZHAN, J., BARRETT, L. G., PAPAIX, J. et THRALL, P. H. (2016). Addressing the challenges of pathogen evolution on the world's arable crops. *Phytopathology*, 106(10):1117–1127.

BURTON, O. J. et TRAVIS, J. M. J. (2008). Landscape structure and boundary effects determine the fate of mutations occurring during range expansions. *Heredity*, 101(4):329–340.

CAROLAN, K., HELPS, J., VAN DEN BERG, F., BAIN, R., PAVELEY, N. et VAN DEN BOSCH, F. (2017). Extending the durability of cultivar resistance by limiting epidemic growth rates. *Proceedings of the Royal Society B : Biological Sciences*, 284.

CORNUET, J. M., RAVIGNÉ, V. et ESTOUP, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, 11(401).

COWGER, C. et BROWN, J. K. M. (2019). Durability of quantitative resistance in crops : Greater than we know ? *Annual Review of Phytopathology*, 57:253–277.

COWGER, C. et MUNDT, C. C. (2002). Effects of wheat cultivar mixtures on epidemic progression of Septoria tritici blotch and pathogenicity of *Mycosphaerella graminicola*. *Phytopathology*, 92(6):617–623.

DANESH, G. (2021). *Phylodynamique des virus évoluant rapidement : approches par calcul bayésien approché et par vraisemblance*. Thèse de doctorat.

de MEEÛS, T. et BALLOUX, F. (2004). Clonal reproduction and linkage disequilibrium in diploids : A simulation study. *Infection, Genetics and Evolution*, 4(4):345–351.

DJIAN-CAPORALINO, C., PALLOIX, A., FAZARI, A., MARTEU, N., BARBARY, A., ABAD, P., SAGE-PALLOIX, A. M., MATEILLE, T., RISSO, S., LANZA, R., TAUSSIG, C. et CASTAGNONE-SERENO, P. (2014). Pyramiding, alternating or mixing : Comparative performances of deployment strategies of nematode resistance genes to promote plant resistance efficiency and durability. *BMC Plant Biology*, 14(1):1–13.

DUPLESSIS, S., CUOMO, C. A., LIN, Y.-C., AERTS, A., TISSERANT, E., VENEAULT-FOURREY, C., JOLY, D. L., HACQUARD, S., AMSELEM, J., CANTAREL, B. L., CHIU, R., COUTINHO, P. M., FEAU, N., FIELD, M., FREY, P., GELHAYE, E., GOLDBERG, J., GRABHERR, M. G., KODIRA, C. D., KOHLER, A., KUES, U., LINDQUIST, E. A., LUCAS, S. M., MAGO, R., MAUCELI, E., MORIN, E., MURAT, C., PANGILINAN, J. L., PARK, R., PEARSON, M., QUESNEVILLE, H., ROUHIER, N., SAKTHIKUMAR, S., SALAMOV, A. A., SCHMUTZ, J., SELLES, B., SHAPIRO, H., TANGUAY, P., TUSKAN, G. A., HENRISSAT, B., VAN DE PEER, Y., ROUZE, P., ELLIS, J. G., DODDS, P. N., SCHEIN, J. E., ZHONG, S., HAMELIN, R. C., GRIGORIEV, I. V., SZABO, L. J. et MARTIN, F. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*, 108(22):9166–9171.

DUPLESSIS, S., LORRAIN, C., PETRE, B., FIGUEROA, M., DODDS, P. N. et AIME, M. C. (2021). Host adaptation and virulence in heteroecious rust fungi. *Annual Review of Phytopathology*, 59:403–422.

DYBIEC, B., KLECZKOWSKI, A. et GILLIGAN, C. A. (2009). Modelling control of epidemics spreading by long-range interactions. *Journal of the Royal Society Interface*, 6(39):941–950.

EDMONDS, C. A., LILLIE, A. S. et CAVALLI-SFORZA, L. L. (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences*, 101(4):975–979.

ESTOUP, A. et GUILLEMAUD, T. (2010). Reconstructing routes of invasion using genetic data : Why, how and so what ? *Molecular Ecology*, 19(19):4113–4130.

EXCOFFIER, L. et RAY, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, 23(7):347–351.

FABRE, B., BASTIEN, C., HUSSON, C., MARÇAIS, B., FREY, P. et HALKETT, F. (2021a). Chapitre 27. Un effet papillon dans les peupleraies françaises : les répercussions d'un contournement de résistance sur les méthodes de sélection variétale. *In* LANNOU, C., ROBY, D., RAVIGNÉ, V., HANNACHI, M. et MOURY, B., éditeurs : *L'immunité des plantes*, pages 329–339. Versailles, quae édition.

FABRE, F., COVILLE, J. et CUNNIFFE, N. J. (2021b). Optimising reactive disease management using spatially explicit models at the landscape scale. *In* SCOTT, P. R., STRANGE, R. N., KORSTEN, L. et GULLINO, M. L., éditeurs : *Plant disease and food security in the 21st century*, pages 47–72. Springer International Publishing.

FABRE, F., ROUSSEAU, E., MAILLERET, L. et MOURY, B. (2015). Epidemiological and evolutionary management of plant resistance : Optimizing the deployment of cultivar mixtures in time and space in agricultural landscapes. *Evolutionary Applications*, 8:919–932.

FAYARD, J., KLEIN, E. et LEFÈVRE, F. (2009). Long distance dispersal and the fate of a gene from the colonization front. *Journal of Evolutionary Biology*, 22(11):2171–2182.

FCBA (2020). Memento 2020. Rapport technique.

FLOR, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, 9(1):275–296.

FREY, P., GÉRARD, P., FEAU, N., HUSSON, C. et PINON, J. (2005). Variability and population biology of *Melampsora* rusts on poplars. *In Rust diseases of willow and poplar*, pages 63–72. CABI Publishing, UK.

GANDON, S. et MICHALAKIS, Y. (2000). Evolution of parasite virulence against qualitative or quantitative host resistance. *Proc. R. Soc. Lond. B*, 267:985–990.

GARNIER, J. (2011). Accelerating solutions in integro-differential equations. *SIAM J. Math. Anal.*, 43:1955–1974.

GÉRARD, P. R., HUSSON, C., PINON, J. et FREY, P. (2006). Comparison of genetic and virulence diversity of *Melampsora larici-populina* populations on wild and cultivated poplar and influence of the alternate host. *Phytopathology*, 96(9):1027–1036.

GILLIGAN, C. A. et van den BOSCH, F. (2008). Epidemiological models for invasion and persistence of pathogens. *Annual Review of Phytopathology*, 46:385–418.

GLADIEUX, P., BYRNES, E. J., AGUILETA, G., FISHER, M. C., HEITMAN, J. et GIRAUD, T. (2011). 4 - Epidemiology and evolution of fungal pathogens in plants and animals. *In* TIBAYRENC, M., éditeur : *Genetics and Evolution of Infectious Diseases*, pages 59–132. Elsevier, London.

GÓMEZ, P., RODRÍGUEZ-HERNÁNDEZ, A. M., MOURY, B. et ARANDA, M. A. (2009). Genetic resistance for the sustainable control of plant virus diseases : breeding, mechanisms and durability. *European Journal of Plant Pathology*, 125(1):1–22.

GOYEAU, H., HALKETT, F., LE CAM, B., MONTARRY, J. et LE MAY, C. (2021). Chapitre 11. Les contournements de résistance : mécanismes, dynamiques et conséquences. *In* LANNOU, C., ROBY, D., RAVIGNÉ, V., HANNACHI, M. et MOURY, B., éditeurs : *L'immunité des plantes*, pages 135–145. Versailles, quae édition.

GRALKA, M., STIEWE, F., FARRELL, F., MÖBIUS, W., WACLAW, B. et HALLATSCHEK, O. (2016). Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters*, 19(8):889–898.

GROSDIDIER, M., IOOS, R., HUSSON, C., CAEL, O., SCORDIA, T. et MARÇAIS, B. (2018). Tracking the invasion : dispersal of *Hymenoscyphus fraxineus* airborne inoculum at different scales. *FEMS microbiology ecology*, 94(5):1–11.

HACQUARD, S., PETRE, B., FREY, P., HECKER, A., ROUHIER, N. et DUPLESSIS, S. (2011). The poplar-poplar rust interaction : Insights from genomics and transcriptomics. *Journal of Pathogens*, 2011:1–11.

HALKETT, F., LANNOU, C., LE MAY, C., MONTARRY, J., PAPAÏX, J. et RAVIGNÉ, V. (2021). Chapitre 10. L'adaptation des agents pathogènes de plantes aux systèmes cultivés : une évolution sous contraintes anthropiques. *In* LANNOU, C., ROBY, D., RAVIGNÉ, V., HANNACHI, M. et MOURY, B., éditeurs : *L'immunité des plantes*, pages 125–134. Versailles, quae édition.

HALLATSCHEK, O., HERSEN, P., RAMANATHAN, S. et NELSON, D. R. (2007). Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930.

HAMELIN, F. M., CASTEL, M., POGGI, S., ANDRIVON, D. et MAILLERET, L. (2011). Seasonality and the evolutionary divergence of plant parasites. *Ecology*, 92(12):2159–2166.

IBRAHIM, K. M., NICHOLS, R. A. et HEWITT, G. M. (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, 77:282–291.

IGN (2022). La production annuelle en volume. *IGN*.

ISLAM, M. T., CROLL, D., GLADIEUX, P., SOANES, D. M., PERSOONS, A., BHATTACHARJEE, P., HOSSAIN, M. S., GUPTA, D. R., RAHMAN, M. M., MAHBOOB, M. G., COOK, N., SALAM, M. U., SUROVY, M. Z., SANCHO, V. B., MACIEL, J. L. N., NHANIJÚNIOR, A., CASTROAGUDÍN, V. L., REGES, J. T. A., CERESINI, P. C., RAVEL, S., KELLNER, R., FOURNIER, E., THARREAU, D., LEBRUN, M. H., MCDONALD, B. A., STITT, T., SWAN, D., TALBOT, N. J., SAUNDERS, D. G. O., WIN, J. et KAMOUN, S. (2016). Emergence of wheat blast in Bangladesh was caused by a South American lineage of Magnaporthe oryzae. *BMC Biology*, 14(1):1–11.

JOHNSON, R. (1979). The concept of durable resistance. *Phytopathology*, 69(3):198.

JOHNSON, R. (1984). A critical analysis of durable resistance. *Annual Review of Phytopathology*, 22.

KINNISON, M. T. et HAIRSTON, N. G. (2007). Eco-evolutionary conservation biology : Contemporary evolution and the dynamics of persistence. *Functional Ecology*, 21(3):444–454.

KLEIN, E., LAVIGNE, C., PICAULT, H., RENARD, M. et GOUYON, P. H. (2006). Pollen dispersal of oilseed rape : Estimation of the dispersal function and effects of field dimension. *Journal of Applied Ecology*, 43(1):141–151.

KLOPFSTEIN, S., CURRAT, M. et EXCOFFIER, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, 23(3):482–490.

KOT, M., LEWIS, M. A. et VAN DEN DRIESSCHE, P. (1996). Dispersal data and spread of invading organisms. 77(7):2027–2042.

LAINE, A. L. et BARRÈS, B. (2013). Epidemiological and evolutionary consequences of life-history trade-offs in pathogens. *Plant Pathology*, 62:96–105.

LE CORRE, V. et KREMER, A. (1998). Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology*, 11: 495–512.

LEYRONAS, C., MORRIS, C. E., CHOUFANY, M. et SOUBEYRAND, S. (2018). Assessing the aerial interconnectivity of distant reservoirs of *Sclerotinia sclerotiorum*. *Frontiers in Microbiology*, 9:1–13.

LO PRESTI, L., LANVER, D., SCHWEIZER, G., TANAKA, S., LIANG, L., TOLLOT, M., ZUCCARO, A., REISSMANN, S. et KAHMANN, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, 66:513–545.

LOF, M. E. et van der WERF, W. (2017). Modelling the effect of gene deployment strategies on durability of plant resistance under selection. *Crop Protection*, 97:10–17.

LOUET, C. (2021). *Analyses évolutive et moléculaire des déterminants génétiques impliqués dans l'interaction entre l'agent de la rouille du peuplier et son hôte*. Thèse de doctorat.

LOUET, C., SAUBIN, M., ANDRIEUX, A., PERSOONS, A., GORSE, M., PÉTROWSKI, J., FABRE, B., DE MITA, S., DUPLESSIS, S., FREY, P. et HALKETT, F. (2021). A point mutation and large deletion at the candidate avirulence locus *AvrMlp7* in the poplar rust fungus correlate with poplar RMlp7 resistance breakdown. *Molecular Ecology*, (February):1–12.

MAILLERET, L. et LEMESLE, V. (2009). A note on semi-discrete modelling in the life sciences. *Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 367(1908):4779–4799.

MARÇAIS, B. et HUSSON, C. (2017). Chalarose du frêne et autres maladies invasives : il est possible de mieux protéger les forêts.

MAUPETIT, A. (2019). *Potentiel évolutif et déterminisme génétique de caractères d'agressivité et morphologiques de l'agent de la rouille du peuplier, Melampsora larici-populina*. Thèse de doctorat.

MCDONALD, B. A. (2004). Population genetics of plant pathogens. *The Plant Health Instructor*.

MCDONALD, B. A. (2010). How can we achieve durable disease resistance in agricultural ecosystems? *New Phytologist*, 185(1):3–5.

MCDONALD, B. A. et LINDE, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, 40:349–379.

MEDLOCK, J. et KOT, M. (2003). Spreading disease : Integro-differential equations old and new. *Math. Biosci.*, 184(2):201–222.

MILGROOM, M. G. (2001). The synthesis of genetics and epidemiology : Contributions of population biology in plan pathology. *Journal of Plant Pathology*, 83(2):57–62.

MILGROOM, M. G. et PEEVER, T. L. (2003). Population biology of plant pathogens : The synthesis of plant disease epidemiology and population genetics. *Plant Disease*, 87(6):608–617.

MILLER, T. E. X., ANGERT, A. L., BROWN, C. D., LEE-YAW, J. A., LEWIS, M., LUTSCHER, F., MARCULIS, N. G., MELBOURNE, B. A., SHAW, A. K., SZŰCS, M., TABARES, O., USUI, T., WEISS-LEHMAN, C. et WILLIAMS, J. L. (2020). Eco-evolutionary dynamics of range expansion. *Ecology*, 101(10):1–14.

MÖLLER, M. et STUKENBROCK, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, 15(12):756–771.

MOUTOU, F. et PASTORET, P. P. (2015). Defining an emerging disease. *Revue Scientifique et Technique de l'OIE*, 34(1):41–44.

MUNDT, C. C. (2002). Use of multiline cultivars and cultivar mixtures for disease management. *Annual Review of Phytopathology*, 40(1):381–410.

MURRAY, J. D. (2002). *Mathematical Biology*, volume 17. Springer-Verlag, third édition.

NATHAN, R., KLEIN, E., ROBLEDO-ARNUNCIO, J. J. et REVILLA, E. (2012). 15 - Dispersal kernels : Review. *In* CLOBERT, J., BAGUETTE, M., BENTON, T. G. et BULLOCK, J. M., éditeurs : *Dispersal ecology and evolution*, pages 186–210. Oxford.

NEI, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3):583–590.

NVSVROT, T., XIA, W., XIAO, Z., ZHAN, C., LIU, M., YANG, X., ZHANG, Y. et WANG, N. (2020). Combining QTL mapping with genome resequencing identifies an indel in an R gene that is associated with variation in leaf rust disease resistance in poplar. *Phytopathology*, 110(4):900–906.

OKUBO, A. et LEVIN, S. A. (2002). *Diffusion and Ecological Problems – Modern Perspectives*. Second edition, Springer-Verlag, New York.

PACILLY, F. C. A., HOFSTEDE, G. J., LAMMERTS VAN BUEREN, E. T., KESSEL, G. J. T. et GROOT, J. C. J. (2018). Simulating crop-disease interactions in agricultural landscapes to analyse the effectiveness of host resistance in disease control : The case of potato late blight. *Ecological Modelling*, 378:1–12.

PARIAUD, B., RAVIGNÉ, V., HALKETT, F., GOYEAU, H., CARLIER, J. et LANNOU, C. (2009). Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathology*, 58(3):409–424.

PARSONS, T. L., LAMBERT, A., DAY, T. et GANDON, S. (2018). Pathogen evolution in finite populations : slow and steady spreads the best. *Journal of royal society*, 15.

PEISCHL, S., KIRKPATRICK, M. et EXCOFFIER, L. (2015). Expansion load and the evolutionary dynamics of a species range. *The American Naturalist*, 185(4):E81–E93.

PERKINS, T. A., PHILLIPS, B. L., BASKETT, M. L. et HASTINGS, A. (2013). Evolution of dispersal and life history interact to drive accelerating spread of an invasive species. *Ecology Letters*, 16(8).

PERSOONS, A. (2015). *Les contournements de résistance par Melampsora larici-populina, l'agent de la rouille du peuplier : impact démographique et déterminisme génétique*. Thèse de doctorat, Université de Lorraine.

PERSOONS, A., HAYDEN, K. J., FABRE, B., FREY, P., DE MITA, S., TELLIER, A. et HALKETT, F. (2017). The escalatory Red Queen : Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, 26(7):1902–1918.

PERSOONS, A., MAUPETIT, A., LOUET, C., ANDRIEUX, A., LIPZEN, A., BARRY, K. W., NA, H., ADAM, C., GRIGORIEV, I. V., SEGURA, V., DUPLESSIS, S., FREY, P., HALKETT, F. et DE MITA, S. (2022). Genomic signatures of a major adaptive event in the pathogenic Fungus *Melampsora larici-populina*. *Genome Biology and Evolution*, 14(1):1–14.

PERSOONS, A., MORIN, E., DELARUELLE, C., PAYEN, T., HALKETT, F., FREY, P., DE MITA, S. et DUPLESSIS, S. (2014). Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. *Frontiers in Plant Science*, 5:1–16.

PETRE, B., HECKER, A., GERMAIN, H., TSAN, P., SKLENAR, J., PELLETIER, G., SÉGUIN, A., DUPLESSIS, S. et ROUHIER, N. (2016). The poplar rust-induced secreted protein (RISP) inhibits the growth of the leaf rust pathogen *Melampsora larici-populina* and triggers cell culture alkalinisation. *Frontiers in Plant Science*, 7(97).

PIETRAVALLE, S., LEMARIÉ, S. et VAN DEN BOSCH, F. (2006). Durability of resistance and cost of virulence. *European Journal of Plant Pathology*, 114:107–116.

PINK, D. et PUDDEPHAT, I. (1999). Deployment of disease resistance genes by plant transformation - a 'mix and match' approach. *Trends in Plant Science*, 4(2):71–75.

PINON, J. et FREY, P. (2005). Interactions between poplar clone and *Melampsora* populations and their implications for breeding for durable resistance. *Rust Diseases of Willow and Poplar*, (July):138–154.

PINON, J., FREY, P., HUSSON, C. et SCHIPFER, A. (1998). Poplar rust (*Melampsora larici-populina*) : the development of E4 pathotypes in France since 1994. *First IUFRO Rusts of Forest Trees*, 712:57–64.

REZNICK, D. N., LOSOS, J. et TRAVIS, J. (2019). From low to high gear : There has been a paradigm shift in our understanding of evolution. *Ecology Letters*, 22(2):233–244.

RIMBAUD, L., FABRE, F., PAPAÏX, J., MOURY, B., LANNOU, C., BARRETT, L. G. et THRALL, P. H. (2021). Models of plant resistance deployment. *Annual Review of Phytopathology*, 59(1).

RIMBAUD, L., PAPAÏX, J., BARRETT, L. G., BURDON, J. J. et THRALL, P. H. (2018). Mosaics, mixtures, rotations or pyramiding : What is the optimal strategy to deploy major gene resistance ? *Evolutionary Applications*, 11:1791–1810.

ROSENBERG, N. A. et NORDBORG, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, 3(5):380–390.

SAPOUKHINA, N., DUREL, C. E. et LE CAM, B. (2009). Spatial deployment of gene-for-gene resistance governs evolution and spread of pathogen populations. *Theoretical Ecology*, 2(4):229–238.

SEVERNS, P. M., SACKETT, K. E., FARBER, D. H. et MUNDT, C. C. (2019). Consequences of long-distance dispersal for epidemic spread : Patterns, scaling, and mitigation. *Plant Disease*, 103(2):177–191.

SHIGESADA, N. et KAWASAKI, K. (1997). *Biological invasions : Theory and practice*. Oxford University Press, UK.

SHINE, R., BROWN, G. P. et PHILLIPS, B. L. (2011). An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences of the United States of America*, 108(14):5708–5711.

SINGH, R. P., HODSON, D. P., JIN, Y., LAGUDAH, E. S., AYLIFFE, M. A., BHAVANI, S., ROUSE, M. N., PRETORIUS, Z. A., SZABO, L. J., HUERTA-ESPINO, J., BASNET, B. R., LAN, C. et HOVMØLLER, M. S. (2015).

Emergence and spread of new races of wheat stem rust fungus : Continued threat to food security and prospects of genetic control. *Phytopathology*, 105(7):872–884.

SLATKIN, M. (1985). Gene flow in natural populations. *Annual review of ecology and systematics*, 16:393–430.

SOMDA, B. et PINON, J. (1981). Ecophysiologie du stade uredien de *Melampsora larici-populina* Kleb. et de *M. allii-populina* Kleb. *European Journal for Pathology*, (11):243–254.

STUKENBROCK, E. H. et MCDONALD, B. A. (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions*, 22(4):371–380.

SUNNAKER, M., BUSETTO, A. G., NUMMINEN, E., CORANDER, J., FOLL, M. et DESSIMOZ, C. (2013). Approximate Bayesian Computation. *PLoS Computational Biology*, 9(1):e1002803.

THRALL, P. H., BARRETT, L. G., DODDS, P. N. et BURDON, J. J. (2016). Epidemiological and evolutionary outcomes in gene-for-gene and matching allele models. *Frontiers in Plant Science*, 6(1084).

TOMA, B. et THIRY, E. (2003). Qu'est ce qu'une maladie émergente ? *Epidémiologie et santé animale*, 44:1–11.

TRAVIS, J. M. J., MÜNKEMÜLLER, T., BURTON, O. J., BEST, A., DYTHAM, C. et JOHST, K. (2007). Deleterious mutations can surf to high densities on the wave front of an expanding population. *Molecular Biology and Evolution*, 24(10):2334–2343.

VAN DEN BOSCH, F. et GILLIGAN, C. A. (2003). Measures of durability of resistance. *Phytopathology*, 93(5): 616–625.

VIALLE, A., FREY, P., HAMBLETON, S., BERNIER, L. et HAMELIN, R. C. (2011). Poplar rust systematics and refinement of *Melampsora* species delineation. *Fungal Diversity*, 50:227–248.

WEI, W. et KRONE, S. M. (2005). Spatial invasion by a mutant pathogen. *Journal of Theoretical Biology*, 236(3):335–348.

WEIR, B. S. et COCKERHAM, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, pages 1358–1370.

WHITLOCK, M. C. et MCCAULEY, D. E. (1999). Indirect measures of gene flow and migration : FST does not equal to 1/(4Nm + 1). *Heredity*, 82:117–125.

WRIGHT, S. (1931). Evolution in mendelian populations. *Genetics*, 16.

WRIGHT, S. (1949). *Adaptation and selection*, chapitre Adaptation, pages 365–389. Princeton Univ. Press, Princeton, NJ.

WRIGHT, S. (1978). *Evolution and the genetics of populations - Variability within and among natural populations*. The University of Chicago Press.

XHAARD, C. (2011). *Influence des processus démographiques sur la structure et les caractéristiques génétiques des champignons phytopathogènes, cas de l'agent de la rouille du peuplier Melampsora larici-populina*. Thèse de doctorat.

XHAARD, C., ANDRIEUX, A., HALKETT, F. et FREY, P. (2009). Characterization of 41 microsatellite loci developed from the genome sequence of the poplar rust fungus, *Melampsora larici-populina*. *Conservation Genetics Resources*, 1(1):21–25.

XHAARD, C., BARRÈS, B., ANDRIEUX, A., BOUSSET, L., HALKETT, F. et FREY, P. (2012). Disentangling the genetic origins of a plant pathogen during disease spread using an original molecular epidemiology approach. *Molecular Ecology*, 21(10):2383–2398.

XHAARD, C., FABRE, B., ANDRIEUX, A., GLADIEUX, P., BARRÈS, B., FREY, P. et HALKETT, F. (2011). The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. *Molecular Ecology*, 20(13):2739–2755.

Annexes

A point mutation and large deletion at the candidate avirulence locus *AvrMlp7* in the poplar rust fungus correlate with poplar RMlp7 resistance breakdown

Clémentine LOUET[1], Méline SAUBIN[1], Axelle ANDRIEUX[1], Antoine PERSOONS[1], Mathilde GORSE[1,2], Jérémy PETROWSKI[1], Bénédicte FABRE[1], Stéphane DE MITA[1,3], Sébastien DUPLESSIS[1], Pascal FREY[1] and Fabien HALKETT[1*]

[1]Université de Lorraine, INRAE, IAM, F-54000 Nancy, France
[2]Present address: INRAE, AgroParisTech, Université Paris-Saclay, BIOGER, F-78850 Thiverval-Grignon, France
[3]Present address: INRAE, Cirad, IRD, Montpellier SupAgro, Université de Montpellier, PHIM, F-34000 Montpellier, France
[*] Corresponding author (fabien.halkett@inrae.fr)

Published in Molecular Ecology, DOI: https://doi.org/10.1111/mec.16294

Abstract

The deployment of plant varieties carrying resistance genes (*R*) exerts strong selection pressure on pathogen populations. Rapidly evolving avirulence genes (*Avr*) allow pathogens to escape *R*-mediated plant immunity through a variety of mechanisms, leading to virulence. The poplar rust fungus *Melampsora larici-populina* is a damaging pathogen of poplars in Europe. It underwent a major adaptive event in 1994, with the breakdown of the poplar *RMlp7* resistance gene. Population genomics studies identified a locus in the genome of *M. larici-populina* that likely corresponds to the candidate avirulence gene *AvrMlp7*. Here, to further characterize this effector, we used a population genetics approach on a comprehensive set of 281 individuals recovered throughout a 28-year period encompassing the resistance breakdown event. Using two dedicated molecular tools, genotyping at the candidate locus highlighted two different alterations of a predominant allele found mainly before the resistance breakdown: a non-synonymous mutation and a complete deletion of this locus. This results in six diploid genotypes: three genotypes related to the avirulent phenotype and three related to the virulent phenotype. The temporal survey of the candidate locus revealed that both alterations were found in association during the resistance breakdown event. They pre-existed before the breakdown in a heterozygous state with the predominant allele cited above. Altogether, these results suggest that the association of both alterations at the candidate locus *AvrMlp7* drove the poplar rust adaptation to *RMlp7*-mediated immunity. This study demonstrates for the first time a case of adaptation from standing genetic variation in rust fungi during a qualitative resistance breakdown.

Keywords

molecular evolution, gene-for-gene interaction, fungal pathogen, effector

1. INTRODUCTION

Fungal plant pathogens are a major threat to food security and biodiversity worldwide (Savary et al. 2019). Therefore, it is essential to gain more insight into the evolutionary history of plant pathogens to better control these harmful microorganisms by knowing the limits of their adaptive potential (Zhan et al., 2015).

To colonize and exploit host plants, fungal pathogens express a specific set of genes encoding secreted proteins (Petre et al., 2020). Among these, pathogenicity factors (or effector proteins) modulate host defense responses. To control diseases, modern agriculture widely uses plants with qualitative resistances relying on major resistance genes (*R*) (Stukenbrock & McDonald, 2008). Typically, R proteins recognize effector proteins and trigger host defense responses (Kourelis & Van Der Hoorn, 2018). The genes encoding such recognized effectors are termed avirulence (*Avr*) genes, according to the gene-for-gene model (Flor, 1971) and the avirulent (Avr) allele is dominant over virulent (avr) allele(s) (Flor, 1956).

R–Avr interactions usually result in an all-or-nothing infection outcome and thus tend to cause strong selection pressure on pathogen populations, ultimately resulting in resistance breakdown (Möller & Stukenbrock, 2017; Scholthof, 2006). *R* genes deployed in genetically uniform monocultures are the main risk factor causing pathogen evolution towards resistance breakdown (McDonald & Linde, 2002). Such breakdowns may result from different types of alteratiæon in *Avr* genes. The two most common alterations are non-synonymous SNPs (single nucleotide polymorphisms) mutating *Avr* genes, and partial or complete deletion of *Avr* genes causing their inactivation (Fouché et al., 2018; Frantzeskakis et al., 2020; Rouxel & Balesdent, 2017; Sánchez-Vallet et al., 2018). A single pathogen species can evolve *via* multiple routes to become virulent on a given host. For instance, in *Leptosphaeria maculans*, virulence towards the same *R*-mediated resistance can be conferred by different alterations at the same locus that coexist in pathogen populations (Daverdin et al., 2012; Rouxel & Balesdent, 2017).

Fundamental questions about the dynamics of alterations during a resistance breakdown remain. Two evolutionary scenarios can be considered. On the one hand, adaptation is driven by a *de novo* mutation: virulent individuals emerge rapidly when a new allele appears. On the other hand, adaptation results from standing genetic variation (Barrett & Schluter, 2008): all alleles already exist in the population, but must be combined properly to produce an adapted phenotype. A way to disentangle these two scenarios is to examine the temporal evolution of the alterations leading to resistance breakdown. Here, we tackled this issue in the poplar rust fungus, a plant pathogen subjected to enduring selection pressure exerted by the deployment of *R* genes in the environment (Xhaard et al., 2011).

The poplar rust fungus *Melampsora larici-populina* (Basidiomycota, Pucciniales) is a biotrophic pathogen. The poplar rust life cycle consists of clonal multiplication on poplars alternating with annual

sexual reproduction on larch (Duplessis et al., 2021). Commercial poplar cultivation is particularly susceptible to poplar rust because of the intensive and monoclonal poplar plantations carrying selected qualitative resistances (Gérard et al., 2006) that are overcome systematically by the pathogen (Pinon & Frey, 2005). The most damaging resistance breakdown occurred in 1994 when the rust resistance gene *RMlp7* (carried by, e.g., *Populus × interamericana* 'Beaupré') was overcome, leading to the invasion in Western Europe of virulent 7 (a7) individuals of *M. larici-populina* (Persoons et al., 2017; Xhaard et al., 2011). Only four years elapsed between detection of the first virulent a7 individuals in Belgium in 1994 and the spread of these individuals throughout France, causing damaging epidemics (Pinon & Frey, 2005; Pinon et al., 1998). Although planting of 'Beaupré' decreased, and indeed totally ceased by 2005, this cultivar still predominated in the landscape for around 20 years, maintaining a strong selection pressure on *M. larici-populina* populations. As a consequence, the genetic structure and distribution of poplar rust populations in France was strongly impacted (Persoons et al., 2017; Xhaard et al., 2011). Three genetic groups were identified in France: (*i*) an ancestral group called 'Fossil', containing only avirulent 7 (A7) individuals, which declined from 1994 and vanished in 1998; (*ii*) a contemporaneous genetic group called 'Cultivated', containing almost exclusively virulent a7 individuals, which emerged at the time of the resistance breakdown (1994), and (*iii*) a third genetic group called 'Wild', containing exclusively avirulent A7 individuals that predominates mostly in south-eastern France where pathogen populations are found mostly on wild poplars and are not affected by commercial poplar plantations (Xhaard et al., 2011; Persoons et al., 2017).

Population genomic analyses identified a unique genomic region under positive selection on chromosome 15 of the *M. larici-populina* 98AG31 reference genome (Persoons et al., 2021; Duplessis et al., 2011). This region harbors a candidate gene that may be associated with *RMlp7*-mediated recognition (Persoons et al., 2021) and thus likely corresponds to the avirulence gene *AvrMlp7*. Three SNPs targeting this gene are associated with the virulence phenotype a7 (Fig. 1), including a non-synonymous SNP at position 2.233.369 (leading to a serine to glycine amino acid substitution at position 81, hereafter abbreviated to $S^{81}G$). Because the A7 and a7 alleles were found predominantly in avirulent A7 or virulent a7 populations, respectively (Persoons et al., 2021), we denominated them as Avr (avirulent) and avr (virulent) alleles. Moreover, 11% of virulent individuals appeared to have no sequence coverage at the candidate locus, which would indicate that this region is deleted in comparison to the 98AG31 reference genome (Persoons et al., 2021). This points to a second putative alteration of the avirulent allele.

The aim of this study was to further elucidate whether evolution of the candidate gene is related to resistance breakdown, and which scenario holds for the evolution of allelic frequencies. For this purpose, we used a historical collection of 281 *M. larici-populina* individuals encompassing the resistance breakdown event (i.e. individuals sampled between 1989 and 2016). We thus expanded to

a greater extent the sampling used in Persoons et al. (2021), which focused on only four populations representative of the three genetic groups. In the following, we (*i*) investigated the occurrence of the two alterations at the candidate locus, (*ii*) assessed genotypes and tested the link with phenotypes, and (*iii*) examined the temporal evolution of allele frequencies in relation to the neutral population structure.

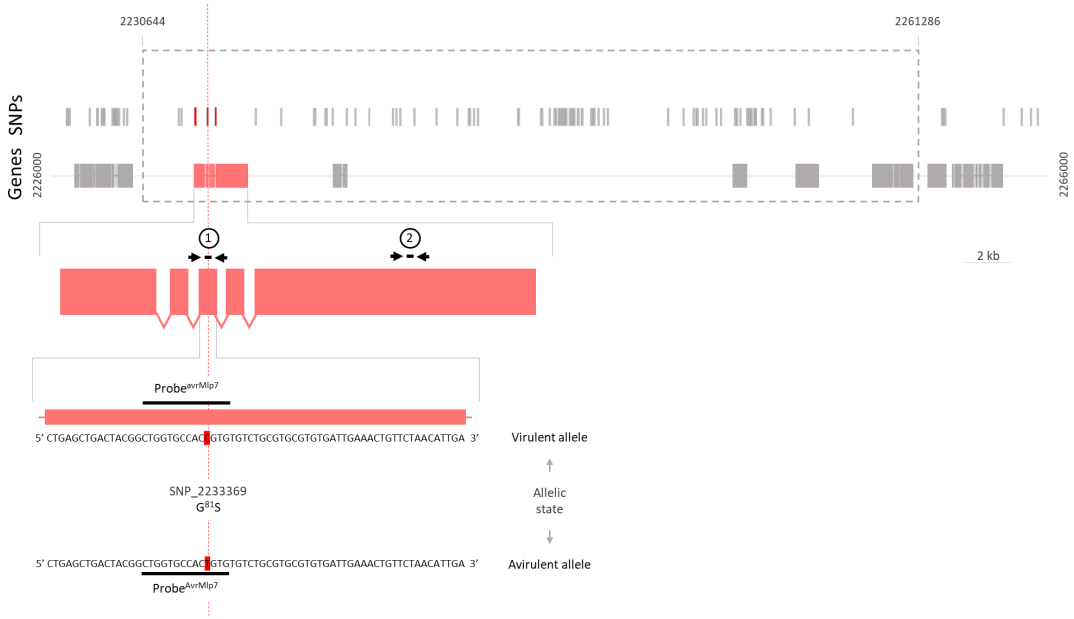


Fig. 1 Genomic organization in the region containing the candidate gene *AvrMlp7*, based on the *Melampsora larici-populina* reference genome 98AG31.

Synonymous single-nucleotide polymorphisms (SNPs) and non-synonymous SNPs are indicated by grey and red bars, respectively. Displayed SNPs occurred in at least 10% of genomes (i.e. minor allele frequency above 0.1). Genes and the candidate gene are indicated in grey and pink, respectively. The exact limits of the deletion are indicated by the grey dotted box. The vertical red dotted line highlights the non-synonymous SNP at position 2.233.369 on chromosome 15 ($C^{241}T$) leading to a $G^{81}S$ amino acid substitution in the predicted protein. Relative positions of primers and probes for the allelic discrimination at the SNP_2233369 (1) and testing the copy number variation (2) are shown in the enlargement of the candidate gene. The third exon nucleotide sequences of the candidate gene and the probe locations (ProbeavrMlp7 and ProbeAvrMlp7) used for allelic discrimination at SNP_2233369 are indicated for both virulent and avirulent alleles.

2. MATERIALS AND METHODS

2.1. *Melampsora larici-populina* individuals

Sampling comprised 354 individuals collected from 1989 to 2016, mainly in the Northeast part of France and Belgium (where the resistance breakdown occurred). These individuals were selected from a historical collection maintained at INRAE Grand-Est Nancy, France. This sampling included most individuals from the previous study of Persoons et al. (2017) and extended the time-period covered by the analysis (Table 1; Table S1, Supporting Information).

Individuals were phenotyped for their ability to overcome *RMlp7*-mediated recognition (Section 2.3) and were genotyped with 22 microsatellite loci (Section 2.4). A total of 73 individuals were excluded from further analyses either because they had missing data, they were clonemates or they appeared to be mixtures of genotypes. Of the remaining 281 individuals, 111 were original for this study.

Table 1 *Melampsora larici-populina* individuals used in the study, with their phenotype towards the *RMlp7*-mediated resistance and their assignment to one of three genetic groups (Fossil, Cultivated, Wild).

	Phenotype[†]			Genetic group			
Year	A7	a7	No. of isolates	Fossil	Cultivated	Wild	Hybrid[‡]
1989	10	-	10	8	-	-	2
1990	5	-	5	5	-	-	-
1991	13	-	13	13	-	-	-
1992	24	-	24	23	-	-	1
1993	19	-	19	19	-	-	-
1994	11	28	39	9	29	1	-
1995	8	22	30	8	22	-	-
1996	3	2	5	1	2	1	1
1997	7	6	13	6	6	1	-
1998	2	22	24	1	21	1	1
1999	3	8	11	-	9	-	2
2001	1	12	13	-	11	2	-
2008	20	30	50	-	30	18	2
2016	2	23	25	-	25	-	-
Total	128	153	281	93	155	24	9

[†]A7, avirulent 7 phenotype towards *RMlp7*; a7, virulent 7 phenotype towards *RMlp7.*
[‡]Hydrids were considered as missing data for subsequent analyses.

2.2. Investigation and delineation of the deletion

The genomes of 66 individuals had previously been sequenced at high depth (80x on average) using Illumina technology, and mapped on the *M. larici-populina* 98AG31 reference genome version 2 (Persoons et al., 2021).

To detect structural variants and identify the boundaries of the deletion, we used GROM (Smith et al., 2017), which operates on paired-end short reads. As recommended by the authors, we aligned reads using the BWA (v.0.7.17) MEM command with default settings. We ran GROM (v.1.0.3) with the -M option (duplicate read filtering). We extracted from GROM's output all structural variants (insertions, deletions or inversions) falling into or overlapping the region located between 2.1 and 2.3 Mbp on chromosome 15. For each variant, we recorded whether these events overlapped with at least one coding exon of any gene, and/or the SNP_2233369, and we computed Pearson's correlation coefficient between the virulence status of each individual and whether it bore this variant.

To examine further the occurrence of the deletion, we computed the sequencing depth at the position of the SNP_2233369 on chromosome 15. In order to correct for differences in sequencing depth among individuals, we calculated the average depth with the surrounding region (2.2–2.3 Mbp). Variation in relative sequencing depth was partitioned using *K*-means clustering as implemented in the R package stats (Cran, v.3.5.2).

2.3. Phenotype assessment

Qualitative virulence phenotypes (i.e. pathotypes) of all individuals were confirmed by inoculation of a poplar cultivar carrying the *RMlp7* resistance gene (*P. x interamericana* 'Beaupré') and of a universal cultivar susceptible to all poplar rust individuals (*P. × euramericana* 'Robusta'), as a positive control, as described in Barrès et al. (2008). Infected 'Robusta' leaf discs of each individual were frozen and stored at –20°C until genomic DNA isolation.

2.4. Genomic DNA isolation and genotyping

Genomic DNA was extracted from infected poplar leaf discs using the DNeasy96 DNA Plant Kit (Qiagen) as described in Maupetit et al. (2018). Genotyping was performed using 22 microsatellite markers (Xhaard et al., 2011) for genotype purity and to avoid potential clones. Genotypes were assigned to the three genetic groups 'Fossil', 'Cultivated' and 'Wild' using the software TESS (Chen et al., 2007) as described in Persoons et al. (2017). Individuals studied in Persoons et al. (2017) were used as controls (Table S1, Supporting Information). A total of 37 individuals displaying mixed membership were removed from further analyses. For population genetic analyses, individuals were pooled according to

sampling location and year, resulting in 14 population samples, with at least eight individuals per sample (mean sample size of 17 individuals).

2.5. Analysis of genotype at the candidate locus

2.5.1. Copy number variation:

To detect deletion of the candidate locus, a TaqMan Copy Number Assay was performed using the software PRIMEREXPRESS (v.3.0.1, Applied Biosystems) to compare the copy number of the gene of interest against a reference gene with known copy number. Ultra-specific qPCR primers for the candidate gene (forward: 5'-CACTAACCCTAACCTAACCTAT-3'; reverse: 5'-TTAAGTGTTGAGTTGCCTAAC-3'; Fig. 1) and for the α-Tubulin gene (JGI protein ID no.73137; forward: 5'-CACTAACCCTAACCTAACCTAT-3'; reverse: 5'-TTAAGTGTTGAGTTGCCTAAC-3) were designed to amplify products of 73 bp and 64 bp, respectively. Two TaqMan fluorogenic probes with different fluorescent dye reporters were used: the first for the candidate gene (5'-FAM-CAACCACCTCGAACCCGCTTCCA-3') and the second for the α-Tubulin gene (5'-VIC-CCCGCTCGCTTCTTACGCCCC-3'). Multiplex assays were performed in technical triplicate in 384-well plates, along with negative controls in each plate, with a final volume of 13 µL containing the following reaction mixture: 2.7 µL of nuclease-free water, 6.5 µL of 2X TaqMan Master Mix (Applied Biosystems), 0.13 µL of each primer at 30 µM, 0.13 µL of each probe at 10 µM and 3 µL of uncalibrated genomic DNA. Quantitative PCR runs were performed on a QuantStudio 6 Flex machine (Life Technologies) with the following cycle parameters: 10 min at 95°C, followed by 50 cycles of 15 s at 90°C and 1 min at 60°C. Assuming a signal intensity proportional to the quantity of gene copies, we considered the number of gene copies as the relative difference between the cycle threshold (Ct) of the candidate gene compared with the Ct of the α-*Tubulin* gene ($2^{-\Delta Ct}$ calculation, Livak & Schmittgen, 2001), present in four copies in the *M. larici-populina* genome (Zhao et al., 2014). Using *K*-means clustering as implemented in the R package STATS (Cran, v.3.5.2), individuals were assigned to three categories depending on the number of gene copies detected: two, one, or none. Attribution to one of these categories was checked using error bars representative of the standard deviation from the mean. If error bars overlapped between categories, individuals were re-tested or excluded from subsequent analyses.

2.5.2. Allelic discrimination at SNP_2233369:

To determine the genotype at SNP_2233369, TaqMan SNP Genotyping Assays were performed using primers and probes designed using PRIMEREXPRESS (v.3.0.1, Applied Biosystems). Two ultra-specific qPCR primers were designed to amplify a 130-bp product encompassing SNP_2233369 of the candidate gene (forward: 5'-ACGTGCTTTTGTGGAGCAGTT-3'; reverse: 5'-TCAGTATGCGTCCTGACTCGTAA-3'). Two fluorogenic TaqMan minor groove binder (MGB) probes were designed, with different fluorescent dye reporters. The FAM-labelled probeAvrMlp7 (5'-FAM-CTGGTGCCACCGTG-MGB-3') and the VIC-labelled probeavrMlp7 (5'-VIC-CTGGTGCCACTGTG-MGB-3') were designed to target specifically the avirulent and virulent alleles of the candidate gene, respectively (Fig. 1). Multiplex assays were performed in technical triplicate in 384-well plates, along with negative controls in each plate, with a final volume of 5 µL containing the following reaction mixture: 2.5 µL of 2X TaqMan Master Mix (Applied Biosystems), 0.25 µL of 20X Assay Working Stock (containing primers and both probes) and 2.25 µL of uncalibrated genomic DNA. Quantitative PCR runs were performed on a QuantStudio 6 Flex machine (Life Technologies) with the following cycle parameters: 30 s at 60°C, 10 min at 95°C, followed by 40 cycles of 15 s at 90°C and 1 min at 60°C and 30 s at 60°C to finish. Fluorescence data were obtained using the QuantStudio Real-Time PCR Software (v.1.3). The average of FAM and VIC fluorescences were normalized by the ROX basal fluorescence present in the 2X TaqMan Master Mix and compared with each other. Using *K*-means clustering as described above, individuals were assigned to four categories: homozygous Avr/Avr, heterozygous Avr/avr, homozygous avr/avr and undetermined. Attribution to one of these categories was checked using error bars representative of the standard deviation from the mean. If error bars overlapped between categories, individuals were re-tested or excluded from subsequent analyses.

2.6. Analysis of genetic diversity at the candidate locus and neutral loci

The genetic diversity of temporal samples was analysed using the nonneutral candidate locus. Hardy-Weinberg equilibrium was tested using Fisher's exact tests as implemented in the R package GENEPOP (Rousset et al., 2020) on all individuals pooled into 14 populations according to the collection year and regardless of the initial location. To estimate the *P*-value for Fisher's exact test, a Markov chain algorithm was implemented using the following parameters: dememorization, 1000; number of batches, 200; iterations per batch, 5000. Using neutral loci (22 microsatellite loci), we also analysed basic population genetics statistics using the R package HIERFSTAT (Goudet & Jombart, 2015) and the population structure with a discriminant analysis of principal components (DAPC, Jombart et al., 2010) as implemented in the R package ADEGENET (Jombart, 2008; Jombart & Ahmed, 2011). In a second step, we grouped individuals according to their genotype at the candidate locus to examine further the link

between neutral genetic variation and the locus under selection. These analyses are presented as supplementary information (Annex S1, Supporting Information).

2.7. Statistical analyses

A Fisher's exact test was used to test the null hypothesis of independence between the genotype at the candidate locus and the phenotype towards *RMlp7*-mediated resistance. This test was performed using the fisher.test R function with a two-sided alternative hypothesis; a *P*-value < 0.001 was considered significant. To evaluate the evolution of allele and genotype frequencies over years, a linear regression was performed using the software R; a *P*-value < 0.05 was considered significant.

3. RESULTS

3.1 A deletion occurs at the candidate locus

With GROM, we detected 339 structural variants in the 2.1–2.3 Mbp region of chromosome 15 using data from Persoons et al. (2021). Among these 339 variants, 202 affected the coding sequence of at least one gene, including 121 deletion or inversion variants affecting SNP_2233369. No insertions were found close to this SNP. Among the 121 variants spanning SNP_2233369, there were 52 large inversions (>1 Mbp and spanning the whole region) unlikely to have a functional effect. Among the remaining 69 deletions, only one displayed a strong positive correlation with virulence a7 phenotype (*r*=0.25). This deletion had a length of 30 kb ranging from position 2.230.644 to 2.261.286 (Fig. 1).

To investigate the occurrence of this deletion at the candidate gene, we examined the variation in sequencing depth at this locus on genomic sequences from 66 whole-genome sequenced individuals. Strong differences were observed (Table S2, Supporting Information). In particular, four individuals showed a sequencing depth lower than 1X both at the SNP_2233369 position and over the flanking region, suggesting the presence of a large deletion estimated between position 2.23 and 2.26 Mbp on chromosome 15 of the 98AG31 reference genome (Fig. 2A; Table S2, Supporting Information). Furthermore, these four individuals (two from 1994 and two from 1998) originated from the 'Cultivated' group and were phenotypically virulent toward the *RMlp7*-mediated resistance. However, the large heterogeneity of sequencing depth at the locus (from 0 to 153.4X) could be indicative of a deletion affecting only one chromosome (i.e. a deletion allele at the heterozygous state). To correct for overall differences in sequencing depth, we computed the average sequencing depth along a larger genomic region ranging from 2.2 to 2.3 Mbp for all other individuals. As illustrated in Fig. 2A, the weighted sequencing depth at the candidate gene *AvrMlp7* clustered in three patterns (for details see Table S2, Supporting Information): a reference pattern exhibiting a regular sequencing depth in comparison with flanking regions, as in the *M. larici-populina* 98AG31 reference genome (pattern A);

a halved sequencing depth in comparison to flanking regions (pattern B); and a lack of reads in this region (pattern C). Altogether, these results identified one unique deletion at the candidate locus affecting either one or both chromosomes.

3.2. The candidate locus displays six possible genotypes

To determine the genotype at the candidate locus, two TaqMan discrimination assays were performed on 281 individuals, after validation on 66 whole-genome sequenced individuals (Table S2, Supporting Information). In a first step, we determined the occurrence of the deletion resulting in the null (Ø) allele. Following the relative signal intensity ($2^{-\Delta Ct}$ ratio), the copy number variation assay indicated that clustering in three categories (*K*max = 3) was the best partition (Fig. 2B). In addition, the three categories identified by the *K*-means partition perfectly matched the grouping into the three sequencing depths described above. Therefore, pattern A corresponds to individuals with two copies of the gene, pattern B corresponds to individuals with one copy of the gene and a null allele, and pattern C corresponds to individuals with two null alleles (denoted as Ø/Ø). In a second step, allelic discrimination at the SNP_2233369 locus was performed to determine the specific alleles at the one or two copies of the gene present: only the avirulent allele, only the virulent allele, or both alleles (i.e. heterozygous individuals) (Fig. S1, Supporting Information). By combining the results of the two assays, six genotypes were identified at the *AvrMlp7* locus: the homozygous genotypes Avr/Avr, avr/avr and Ø/Ø and the heterozygous genotypes Avr/avr, Avr/Ø and avr/Ø. No significant departure from Hardy-Weinberg proportions was observed (*P*-values ranging from 0.11 to 1; Table S3, Supporting Information). To conclude, the three alleles can be found in random association at the genotype level.

3.3. Genotypes at the candidate locus are linked to virulence phenotypes

To evaluate whether the candidate avirulence gene is involved in *RMlp7*-mediated recognition, we compared the genotypic states at the candidate locus with the qualitative virulence phenotype for the

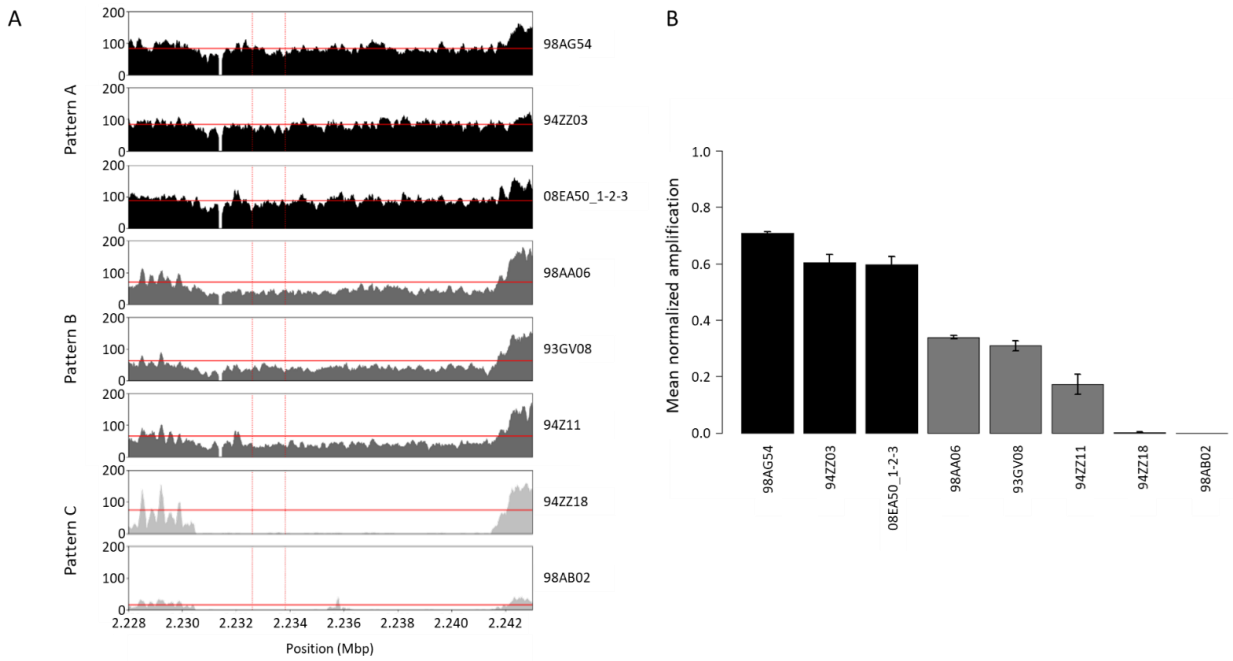


Fig. 2 Patterns of sequencing depth in the region containing the candidate locus *AvrMlp7* and amplification of the candidate gene in eight *Melampsora larici-populina* individuals.

(A) Raw sequencing depth from eight individuals after mapping of the sequencing reads on chromosome 15 of the 98AG31 reference genome. The region impacted by the deletion was shortened to better illustrate the three distinct patterns of sequencing depth among individuals: pattern A with a regular sequencing depth in comparison with flanking regions, pattern B with a decreased sequencing depth in comparison with flanking regions and pattern C with a total lack of sequencing depth in comparison with flanking regions. The vertical red dotted lines encompass the exact position of the candidate gene and the horizontal red line represents the average depth between flanking regions of the candidate locus (2.2 to 2.3 Mbp) along chromosome 15. (B) Mean normalized amplification of the candidate gene from the same eight individuals. The colors represent the individual assignment by *K*-means clustering (*K*max = 3). Error bars indicate the standard deviation of the mean.

281 *M. larici-populina* individuals (Table 2; Table S1, Supporting Information). Overall, three genotypes (Avr/Avr, Avr/avr and Avr/Ø) are associated with avirulent A7 individuals and the other three genotypes (avr/avr, avr/Ø and Ø/Ø) are associated with virulent a7 individuals (Table 2), suggesting dominance of the avirulent allele and a gene-for-gene relationship scenario. Only five individuals (out of 281), i.e. 1.8% of analyzed samples, presented a discrepancy between the genotype and the expected phenotype according to the gene-for-gene relationship. The match between genotypes and expected virulence phenotypes was highly significant according to a Fisher's exact test; the probability of observing this result by chance was $P < 0.001$ (P-value = 0.0005). In conclusion, we observed a tight link between the genotype at the candidate avirulence locus and the virulence phenotype.

Table 2 Genotypes identified at the candidate locus *AvrMlp7* in relation to phenotypes towards the *RMlp7* resistance gene.

		Genotype‡						
Phenotype†		Avr / Avr	Avr / avr	Avr / Ø	avr / avr	avr / Ø	Ø / Ø	Total
	A7	64	27	33	-	1	3	128
	a7	-	1	-	44	58	50	153
	Total	64	28	33	44	59	53	281

†A7, avirulent 7 phenotype towards *RMlp7*; a7, virulent 7 phenotype towards *RMlp7.*
‡ Genotype at the *AvrMlp7*

3.4. Both the virulent and the null alleles expanded rapidly at the time of resistance breakdown

To study the temporal evolution of the polymorphism at the candidate locus during the RMlp7 resistance breakdown event, we calculated allelic frequencies according to years and genetic groups (Fig. 3). In the 'Fossil' group, including only avirulent A7 individuals, we observed a predominance of the avirulent allele (75%) with no significant difference in frequency over time (*P*-value = 0.610). Notably, both the virulent and the null alleles were identified in the 'Fossil' group (8% and 17%, respectively), already five years before the breakdown of *RMlp7*-mediated resistance. In the 'Wild' group containing only avirulent A7 individuals, we also observed the predominance of the avirulent allele (69%). Conversely, in the 'Cultivated' group, emerging at the time of the resistance breakdown in 1994, the avirulent allele had almost disappeared (0.01%) in favor of the virulent (48%) and the null (51%) alleles. In addition, the frequency of the null allele increased significantly (*P*-value = 0.019) from 17% in 1994 to 82% in 2016, whereas the frequency of the virulent allele decreased significantly during this period (*P*-value = 0.012). Therefore, the allelic evolution of the candidate locus coincided with the change in neutral population structure of *M. larici-populina* (Annex S1, Supporting Information). To conclude, the RMlp7 resistance breakdown event is associated with a concomitant increase in frequency of both the virulent and the null alleles, and the null allele has taken over in recent years.

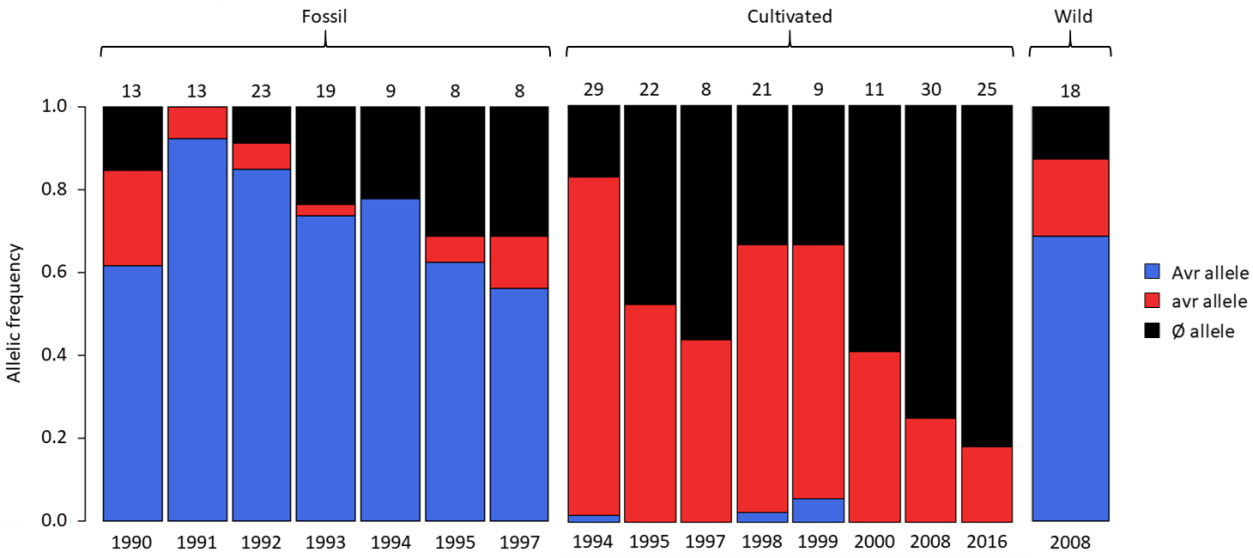


Fig. 3 Cumulative frequencies of alleles at the candidate locus *AvrMlp7* in *Melampsora larici-populina* individuals according to genetic group and sampling years.

Numbers at the top indicate the number of individuals assayed. To balance sample sizes, the years 1990 and 1997 in Fossil group gather individuals collected in 1989–1990 and 1996–1997–1998, respectively and the year 1997 in Cultivated group gather individuals collected in 1996–1997. Avr: avirulent allele; avr: virulent allele; Ø: null allele.

4. DISCUSSION

Our study has helped elucidate the complexity of a double molecular alteration in a diploid organism, the poplar rust fungus, *M. larici-populina*. We based our analyses on a comprehensive temporal sampling that was made possible thanks to a large historical collection covering 28 years of contemporaneous evolution encompassing the RMlp7 breakdown event. Three alleles were detected in a specific candidate for the avirulence gene *AvrMlp7*, the avirulent allele that predominated initially, and the virulent and null alleles that burst onto the scene at the time of resistance breakdown. The evolution of this candidate gene is related to the emergence of the virulent phenotype a7 and the change in genetic structure in poplar rust populations.

4.1. The candidate gene seems to follow a gene-for-gene relationship

Through our screening of 281 individuals of *M. larici-populina*, we found an almost perfect concordance between genotypes at the candidate locus and phenotypes towards the *RMlp7*-mediated resistance. Three genotypes (Avr/Avr, Avr/avr and Avr/Ø) are associated with avirulence and the three others (avr/avr, avr/Ø and Ø/Ø) with virulence. Assuming a gene-for-gene relationship (Flor, 1971), our results are consistent with the avirulent allele being dominant over other alterations conferring virulence, namely the virulent and the null alleles. However, we found five individuals out of the 281

studied that did not follow this phenotype-genotype match, suggesting the existence of other alterations not yet identified. Our analysis of candidate gene polymorphisms is based on the genomic variation identified from a limited number of genomes (Persoons et al. 2021). By enlarging the screening of individuals, we broadened the genomic background, hence revealing mismatches that would be indicative of other putative alterations occurring at a much-reduced frequency. A given avirulence gene can be altered in different ways (Rouxel & Balesdent, 2017). These additional genomic alterations may impact pathogen fitness more heavily, which could explain their reduced frequency, as observed for viruses (Fabre et al. 2012). Conversely, the two alterations identified here would be the least costly, and were therefore retained throughout the evolution of poplar rust populations during resistance breakdown.

Until now, only a handful of *Avr* genes have been characterized in rust fungi (Chen et al., 2017; Petre et al., 2014; Salcedo et al., 2017; Upadhyaya et al., 2021; Wu et al., 2020). The obligate biotrophy nature of rust fungi represents a challenge, which explains the limited number of functional studies (Lorrain et al., 2019). Some authors have managed to bypass these limitations with indirect characterization using progenies and genetic mapping, e.g. in the wheat stem rust *Puccinia graminis* f. sp. *tritici* (Zambino et al., 2000), in the pine rust *Cronartium quercuum* f. sp. *fusiforme* (Kubisiak et al., 2011) and, most recently, in the wheat yellow rust *Puccinia striiformis* f. sp. *tritici* (Wang et al., 2018). Our study proved that population genetics can be an efficient method with which to identify and investigate a putative avirulence gene in diploid pathogens. A formal demonstration of avirulence function will require functional approaches targeting the avirulent allele.

4.2. Virulence towards RMlp7 results from two non-exclusive molecular mechanisms

Here, we identified two non-exclusive alterations at the candidate locus *AvrMlp7* within diploid genotypes. Generally, an allele resulting from non-synonymous SNP(s) presents an altered version of the avirulence Avr protein, allowing escape from recognition mediated by the corresponding *R* protein (Ve et al., 2013; Wang et al., 2007). Nevertheless, some virulent pathogens are characterized mainly by knockout alterations such as the complete loss of the avirulence gene. In the wheat stem rust fungus *P. graminis* f. sp. *tritici*, the genes *AvrSr35* and *AvrSr50* contain large insertions of 57 kbp and 26 kbp, respectively, that resulted from transposon-mediated disruptions (Chen et al., 2017; Salcedo et al., 2017). Similarly, a deletion of a large genomic region containing an avirulence gene was shown to be responsible for a gain in virulence in the wheat pathogen *Zymoseptoria tritici* (Hartmann et al., 2017) and the oilseed rape pathogen *L. maculans* (Gout et al., 2007). In our study, sequencing depth analyses highlighted that the two breakpoints of the deletion encompassing the candidate gene *AvrMlp7*

exhibited the same motif, suggesting a single deletion event, which then spread widely within poplar rust populations.

In haploids, any alteration of the avirulence gene (e.g. a non-synonymous mutation (avr) or gene inactivation caused by a deletion (Ø)) determines the virulence phenotype in a straightforward relationship (Fig. 4A, B). For diploid pathogens, the relationship is more complex as alleles conferring virulence are usually recessive (Flor, 1956). In this case, the occurrence of an Avr allele (even in a heterozygous state) is sufficient to trigger plant resistance (Fig. 4C). It was thus already shown that a diploid pathogen can evolve through several alteration events leading to virulence towards an *R* gene. High levels of polymorphism in avirulence genes have also been reported in other rust fungi, as in the flax rust fungus *Melampsora lini,* where proteins AvrL567 and AvrM exhibit multiple allelic variants or a copy number variation with different alleles, directly impacting pathogenicity (Catanzariti et al., 2006; Dodds et al., 2004). Here, we demonstrated that any association of alleles can confer the virulence phenotype (Fig. 4D).

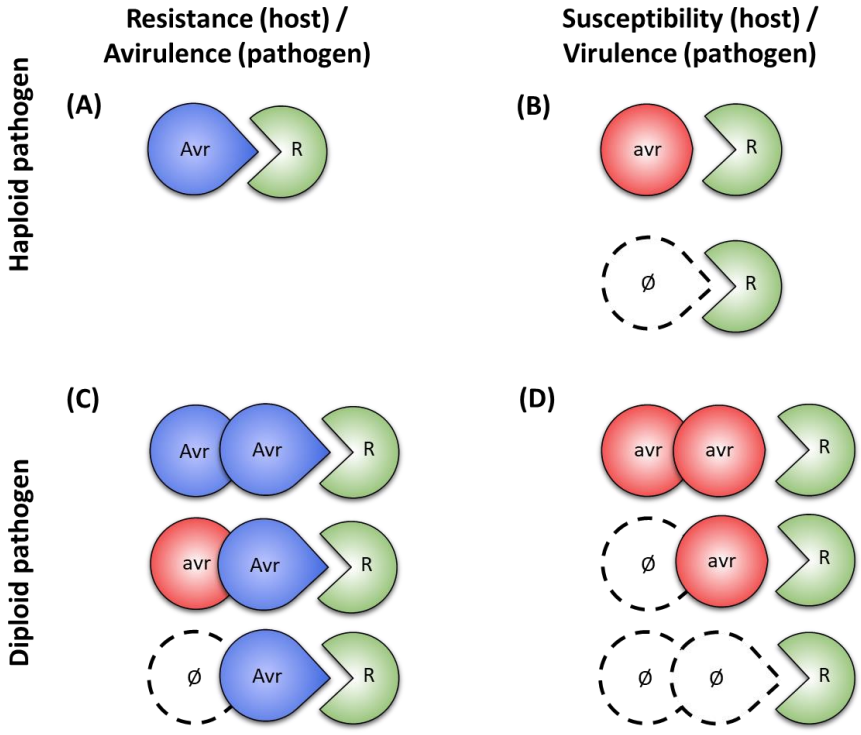


Fig. 4 Difference in molecular determinism of R-mediated resistance between haploid and diploid pathogens.

For haploid pathogens, the outcome of the interaction between the pathogen's avirulence gene and the resistance gene (*R*) is straightforward: (A) the effector protein encoded by the avirulent allele (Avr) is recognized, directly or indirectly, by the *R* gene product, which triggers host resistance; (B) the virulent allele, e.g. a non-synonymous mutation (avr) or a deletion (Ø) of the avirulent allele leads to host susceptibility because of a lack of recognition by the *R* gene product. For diploid pathogens the interaction is complicated by the dominance of the avirulent allele: (C) a single haploid copy of the avirulent allele is sufficient for recognition by the *R* gene product and initiation of host resistance; (D) only the association of two recessive virulent alleles can lead to host susceptibility.

4.3. A point mutation and a large deletion were combined during resistance breakdown

Our study revealed the presence of the two alterations at least five years before RMlp7 resistance breakdown, but, once this resistance was overcome, their frequency increased rapidly. Therefore, adaptation of *M. larici-populina* to *RMlp7*-mediated resistance resulted from selection from standing genetic variation (Barrett & Schluter, 2008). However, at least five years elapsed before the establishment of a favorable combination leading to virulence towards the *RMlp7*-mediated resistance. The dikaryotic state of rust fungi impedes the evolution of virulence, because of the dominance of the avirulent allele. Alleles conferring virulence are subjected mostly to drift as long as they are in the heterozygous state. In this respect, a recent modelling study on the dynamics of resistance breakdown highlighted that the diploid state leads to much more unpredictable evolutionary trajectories, with a delayed time for fixation of the virulence compared with haploids (Saubin et al., 2021). This phenomenon is further marked in pathogens such as *M. larici-populina* that require an alternate host to complete their life cycle. Indeed, obligate migration to resistant and susceptible hosts after sexual reproduction on the alternate host causes high rates of mortality of avirulent individuals, especially when the resistant host predominates in the landscape. As long as virulent allele(s) frequency is low, no genotypes missing the avirulent allele are produced, hence no phenotypically virulent individual exists. By contrast, once the resistance is overcome, the virulent allele(s) spread massively in the landscape (over both resistant and susceptible hosts) and the avirulent allele collapses rapidly, even in individuals living on susceptible hosts (Saubin et al., 2021). Such modeling work is in accordance with our observations. Here, the scenario is complicated by the occurrence of the virulent and null alleles, conferring virulence. The fact that they are not mutually exclusive likely fostered the pace of this breakdown event and contributed to the rapid emergence of the 'Cultivated' group. According to this scenario, the 'Cultivated' population might have been founded from a single event, as previously suggested (Persoons et al., 2017). Replacement of the 'Fossil' population (described in Persoons et al., 2017) coincided with the replacement of the avirulent allele by the two alleles conferring virulence.

4.4. Does the null allele confer a fitness advantage?

Whereas the two alleles conferring virulence exploded at the time of resistance breakdown, their relative frequency was not stable through time afterwards. We observed a significant increase in the frequency of one allele leading to the predominance of the null allele in contemporary populations. This variation can be neutral (resulting from genetic drift), but may also suggest a consequence on fitness for individuals carrying either allele. Genetic modifications leading to the acquisition of a new virulence may be accompanied by unfavorable effects (Leach et al., 2001). Indeed, the avirulence genes

are involved primarily in various functions necessary for infection (Khan et al., 2018). Any alteration at the avirulence locus can thus translate into a reduction in fitness, the so-called fitness cost of virulence (van der Plank, 1963). No such cost has been reported so far in poplar rust; virulent individuals were found to display the same level of disease-associated traits as avirulent individuals (Maupetit et al., 2021). However, this latter study did not consider the two types of alterations conferring virulence. The deletion may provide an advantage over point mutation mechanisms because of the absence of cost associated with the synthesis of the mutated protein. Examining the fitness advantage or disadvantage of each allele in more detail will require a dedicated study, monitoring several traits associated with completion of the infection process and life cycle.

5. CONCLUSION

Efforts dedicated to developing new analysis strategies for avirulence genes have focused mostly on haploid organisms (Bakkeren & Szabo, 2020). Here, we proposed a method combining population genetics and molecular approaches to investigate the candidate avirulence gene *AvrMlp7* in *M. larici-populina*, a diploid fungus responsible for poplar rust disease. We identified two alterations at the candidate locus: a single nucleotide mutation and a complete deletion. These two alleles co-occurred within pathogen populations and associated randomly within individuals. To our knowledge, this is the first demonstration that two alterations conferring virulence can combine within a diploid individual. Their association increased the chance of observing virulent phenotypes and was probably responsible for the breakdown of *RMlp7*-mediated resistance in 1994.

ACKNOWLEDGMENTS

We thank Jean Pinon for creating the historical poplar rust collection at INRAE Nancy in the 1980s. We warmly thank Fanny Hartmann for constructive comments on a previous version of the manuscript, and Benjamin Petre for proofreading an early version of manuscript, a thousand blessings to him. We also thank Tatiana Giraud and four anonymous reviewers for detailed comments and suggestions which helped improve the manuscript. Helen Rothnie corrected the English language. This work was supported by grants from the French National Research Agency (ANR-18-CE32-0001, Clonix2D project; ANR-17-CE32-0010, WABSARF project; ANR-11-LABX-0002-01, Cluster of Excellence ARBRE). Clémentine Louet was supported by a PhD fellowship from the Région Lorraine and the French National Research Agency (ANR-18-CE32-0001, Clonix2D project). Méline Saubin was supported by a PhD fellowship from INRAE and the French National Research Agency (ANR-18-CE32-0001, Clonix2D project). Antoine Persoons was supported by a PostDoc fellowship from the Region Grand-Est and the French National Research Agency (ANR-11-LABX-0002-01, Lab of Excellence ARBRE).

DATA ACCESSIBILITY

Virulence phenotype towards the *RMlp7-mediated* resistance and genotype at the *AvrMlp7* locus as well as microsatellite data from the 281 individuals will be made available on Dryad repository and released at the time of publication.

AUTHOR CONTRIBUTIONS

C.L., M.S., P.F. and F.H. conceived and designed the study. C.L. and A.A. designed and performed molecular experiments. M.S., J.P. and B.F. performed individual cultures and virulence type analyses. C.L., M.S., M.G. and F.H. performed the genotyping. S.D.M. and F.H. performed the alignment. C.L., M.S., S.D.M., A.P. and F.H. analyzed the data. C.L. and F.H. prepared the manuscript. All authors revised and approved the manuscript.

REFERENCES

Bakkeren G, Szabo LJ (2020) Progress on molecular genetics and manipulation of rust fungi. *Phytopathology*, **110**, 532–543

Barrès B, Halkett F, Dutech C *et al.* (2008) Genetic structure of the poplar rust fungus *Melampsora larici-populina*: evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, **8**, 577–587.

Barrett LG, Thrall, PH, Burdon JJ, Linde CC (2008). Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends in Ecology and Evolution*, **23**, 678–685.

Catanzariti AM, Dodds PN, Lawrence GJ, Ayliffe MA, Ellis, JG (2006) Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *The Plant Cell*, **18**, 243–256.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.

Chen J, Upadhyaya NM, Ortiz D *et al.* (2017) Loss of *AvrSr50* by somatic exchange in stem rust leads to virulence for *Sr50* resistance in wheat *Science*, **358**, 1607–1610.

Daverdin G, Rouxel T, Gout L *et al.* (2012) Genome structure and reproductive behaviour influence the evolutionary potential of a fungal phytopathogen. *PLoS Pathogens*, **8**, e1003020.

Dodds PN, Lawrence GJ, Catanzariti AM, Ayliffe MA, Ellis JG (2004) The *Melampsora lini AvrL567* avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *The Plant Cell*, **16**, 755–768.

Duplessis S, Lorrain C, Petre B *et al.* (2021) Host adaptation and virulence in heteroecious rust fungi. *Annual Review of Phytopathology*, **59**, 171-1720.

Flor HH (1956) The complementary genic systems in flax and flax rust. *Advances in Genetics*, **8**, 29–54.

Flor HH (1971) Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, **9**, 275–296.

Fouché S, Plissonneau C, Croll D (2018) The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Current Opinion in Microbiology*, **46**, 34–42.

Frantzeskakis L, Di Pietro A, Rep M, *et al.* (2020) Rapid evolution in plant–microbe interactions – a molecular genomics perspective. *New Phytologist*, **225**, 1134–1142.

Gérard PR, Husson C, Pinon J, Frey P (2006) Comparison of genetic and virulence diversity of *Melampsora larici-populina* populations on wild and cultivated poplar and influence of the alternate host. *Phytopathology*, **96**, 1027–1036.

Goudet J, Jombart T (2015) Hierfstat: estimation and tests of hierarchical F-statistics.

Gout L, Kuhn ML, Vincenot L *et al.* (2007) Genome structure impacts molecular evolution at the *AvrLm1* avirulence locus of the plant pathogen *Leptosphaeria maculans. Environmental Microbiology*, **9**, 2978–2992.

Hacquard S, Petre B, Frey P *et al.* (2011) The poplar-poplar rust interaction: insights from genomics and

transcriptomics. *Journal of Pathogens*, **2011**.
Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D (2017) A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *The ISME Journal*, **11**, 1189–1204.
Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
Jombart T, Ahmed I (2011) adegenet 13-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.
Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.
Khan M, Seto D, Subramaniam R, Desveaux D (2018) Oh, the places they'll go! A survey of phytopathogen effectors and their host targets. *The Plant Journal*, **93**, 651–663.
Kourelis J, Van Der Hoorn RAL (2018) Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *The Plant Cell*, **30**, 285–299.
Kubisiak TL, Anderson CL, Amerson HV *et al.* (2011) A genomic map enriched for markers linked to *Avr1* in *Cronartium quercuum* fsp *fusiforme. Fungal Genetics and Biology*, **48**, 266–274.
Leach JE, Vera Cruz CM, Bai J, Leung H (2001) Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annual Review of Phytopathology*, **39**, 187–224.
Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*, **25**, 402–408.
Lorrain C, Gonçalves dos Santos KC, Germain H, Hecker A, Duplessis S (2019) Advances in understanding obligate biotrophy in rust fungi. *New Phytologist*, **222**, 1190–1206.
Maupetit A, Larbat R, Pernaci M *et al.* (2018) Defense compounds rather than nutrient availability shape aggressiveness trait variation along a leaf maturity gradient in a biotrophic plant pathogen. *Frontiers in Plant Science*, **9**, 1396.
Maupetit A, Fabre B, Pétrowski J *et al.* (2021) Evolution of morphological but not aggressiveness-related traits following a major resistance breakdown in the poplar rust fungus, *Melampsora larici-populina*. *Evolutionary Applications*, **14**, 513–523.
McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, **40**, 349–379.
Möller M, Stukenbrock EH (2017) Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, **15**, 756–771.
Persoons A, Hayden KJ, Fabre B *et al.* (2017) The escalatory Red Queen: population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, **26**, 1902–1918.
Persoons A, Maupetit A, Louet C *et al.* (2021) Genomic signatures of a major adaptive event in the pathogenic fungus *Melampsora larici-populina. BioRxiv* 2021.04.09.439223.
Petre B, Joly DL, Duplessis, S (2014) Effector proteins of rust fungi. *Frontiers in Plant Science*, **5**, 416.
Petre B, Lorrain C, Stukenbrock EH, Duplessis S (2020) Host-specialized transcriptome of plant-associated organisms. *Current Opinion in Plant Biology*, **56**, 81–88.
Pinon J, Frey P (2005) Interactions between poplar clones and *Melampsora* populations and their implications for breeding for durable resistance. In: *Rust diseases of willow and poplar* (eds Pei MH, McCraken AR), pp. 139–154. CAB International, Wallingford, UK.
Pinon J, Frey P, Husson C, Schipfer A (1998) Poplar rust (*Melampsora larici-populina*): the development of E4 pathotypes in France since 1994. In: *Proceedings of the First IUFRO Rusts of Forest Trees Conference*, Saariselkä, Finland (eds Jalkanen R, Crane PE, Walla JA, Aalto T), pp. 57–64. Finnish Forest Research Institute, Research Paper 712, Saarijïarvi.
Rousset F, Lopez J, Belkhir K (2020) genepop: population genetic data analysis using genepop.
Rouxel T, Balesdent MH (2017) Life, death and rebirth of avirulence effectors in a fungal pathogen of *Brassica crops, Leptosphaeria maculans. New Phytologist*, **214**, 526–532.
Salcedo A, Rutter W, Wang S *et al.* (2017) Variation in the *AvrSr35* gene determines *Sr35* resistance against wheat stem rust race Ug99. *Science*, **358**, 1604–1606.
Sánchez-Vallet A, Fouché S, Fudal I *et al.* (2018) The genome biology of effector gene evolution in filamentous plant pathogens. *Annual Review of Phytopathology*, **56**, 21-210.
Saubin M, De Mita S, Zhu X, Sudret B, Halkett F (2021) Impact of ploidy and pathogen life cycle on resistance durability. *BioRxiv* 2021.05.28.446112.
Savary S, Willocquet L, Pethybridge SJ *et al.* (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology and Evolution*, **3**, 430–439.
Scholthof KBG (2006) The disease triangle: pathogens, the environment and society. *Nature Reviews*

Microbiology, **5**, 152–156.
Smith SD, Kawash JK, Grigoriev A (2017) Lightning-fast genome variant detection with GROM. *GigaScience*, **6**, 1–7.
Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, **46**, 75–100.
Upadhyaya NM, Mago R, Panwar V *et al.* (2021) Genomics accelerated isolation of a new stem rust avirulence gene-wheat resistance gene pair. *Nature Plants*, **7**, 1220–1228.
van der Plank JE (1963). *Plant diseases: Epidemics and control*. Academic Press, New York, 349 p.
Ve T, Williams SJ, Catanzariti AM *et al.* (2013) Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *PNAS*, **110**, 17594–17599.
Wang CIA, Guncar G, Forwood JK *et al.* (2007) Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *The Plant Cell*, **19**, 2898–2912.
Wang L, Zheng D, Zuo S *et al.* (2018) Inheritance and linkage of virulence genes in chinese predominant race CYR32 of the wheat stripe rust pathogen *Puccinia striiformis* f sp *tritici. Frontiers in Plant Science*, **9**, 120.
Wu JQ, Dong C, Song L, Park RF (2020) Long-read–based de novo genome assembly and comparative genomics of the wheat leaf rust pathogen *Puccinia triticina* identifies candidates for three avirulence genes. *Frontiers in Genetics*, **11**, 521.
Xhaard C, Fabre B, Andrieux A *et al.* (2011) The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. *Molecular Ecology*, **20**, 2739–2755.
Xu X (2012) Super-races are not likely to dominate a fungal population within a life time of a perennial crop plantation of cultivar mixtures: a simulation study. *BMC Ecology*, **12**, 1–10.
Zambino PJ, Kubelik AR, Szabo LJ (2000) Gene action and linkage of avirulence genes to DNA markers in the rust fungus *Puccinia graminis. Phytopathology*, **90**, 819–826.
Zhan J, Thrall PH, Papaïx J, Xie L, Burdon JJ (2015) Playing on a pathogen's weakness: using evolution to guide sustainable plant disease control strategies. *Annual Review of Phytopathology*, **53**, 19–43.
Zhao Z, Liu H, Luo Y *et al.* (2014) Molecular evolution and functional divergence of tubulin superfamily in the fungal tree of life. *Scientific Reports*, **4**, 6746.

SUPPORTING INFORMATION

Annex S1 Analysis of the neutral genetic diversity and population structure of temporal samples collected in Amance (North-East France).

Fig. S1 Illustration of allelic discrimination of the SNP_2233369 at the *AvrMlp7* gene for a subset of *Melampsora larici-populina* whole-genome sequenced individuals.

Table S1 List and characteristics of *Melampsora larici-populina* individuals used in this study.

Table S2 Polymorphism identification at the *AvrMlp7* locus in the 66 whole-genome sequenced individuals of *Melampsora larici-populina*.

Table S3 Analysis of Hardy-Weinberg equilibrium for the non-neutral SNP_2233369 locus of *Melampsora larici-populina* temporal samples from Amance, France.

Title: Influence of demographic variations on the genetic structure of populations in the context of emerging disease

Keywords: Demogenetics; Population genetics; Epidemiology; Modelling; Phytopathology

Abstract: The demography and genetic structure of a population are closely linked. The study of this interplay is crucial, especially for organisms with frequent demographic fluctuations such as pathogen species responsible for emerging diseases. Classical population genetics models have been used to explore this link for simplified demographic processes. In this thesis, I further investigate the interplay between demography and genetic evolution in pathogen species that display complex life cycles. For this purpose, I focus on two ecological systems that strongly deviate from demographic equilibrium, each of which allows for realistic modelling assumptions. The first ecological system focuses on a major selection event with the temporal dimension being of prime importance. The second system reports recurrent colonisation events in which the spatial aspect is decisive.

In the first part, I study an event of resistance overcoming by a pathogen population. Firstly, a modelling approach is used to establish the conditions for observing resistance overcoming and to identify the determinants of resistance durability. The results highlight the antagonistic effect of the proportion of resistant hosts deployed in the agricultural landscape, which decreases the probability of overcoming but increases the speed of overcoming when it occurs. Secondly, this model is implemented to account for genetic evolution at neutral loci. The results identify three demographic scenarios associated with distinct genetic signatures during resistance overcoming: 1) Small variations in population sizes and small changes in genetic structures, 2) A strong founder event on the resistant host that in turn creates a genetic structure on the susceptible host, and 3) An evolutionary rescue event that results in a strong founder event on the resistant host, preceded by a bottleneck on the susceptible host. Finally, this theoretical framework of demogenetic analysis is applied to empirical data to infer the parameters underlying the overcoming of resistant RMlp7 poplars by the pathogen *Melampsora larici-populina*. Two parameters are particularly well estimated and the inferred values are in agreement with our biological knowledge: a high proportion of resistant hosts in the landscape (more than 80%) and an initial frequency of virulent alleles in the pathogen population between 5 and 10%.

In the second part, I study colonisation and its genetic consequences. These analyses focus on the recurrent invasion of the Durance River valley by *Melampsora larici-populina*. Firstly, a mechanistic-statistical model is coupled to epidemiological data to infer the parameters underlying the pathogen's expansion dynamics. This approach shows that colonisation results from frequent long-distance dispersal events, with an average dispersal distance of more than two kilometres. Secondly, the characterisation of several annual colonisation events highlights a similar genetic structure which amplitude however varies greatly between years. Two extremes are identified: in 2011, strong conservation of the initial genetic diversity along the colonised domain; in 2004, rapid erosion of genetic diversity. The inter-annual variations in these structures can be explained by variations in the frequency of long-distance dispersal events.

This work highlights the importance of contextualised models that take into account demogenetic variations for a better understanding of biological systems. The information obtained is then crucial for developing relevant control strategies against pathogen populations responsible for emerging diseases.

Titre : Effets des variations démographiques sur la structure génétique de populations, dans le cadre d'une maladie émergente

Mots-clés : Démogénétique ; Génétique des populations ; Épidémiologie ; Modélisation ; Phytopathologie

Résumé : La démographie et la structure génétique d'une population sont étroitement liées. L'étude de ce lien est cruciale, ce d'autant plus pour les organismes présentant des fluctuations démographiques fréquentes tels que les agents pathogènes responsables de maladies émergentes. Les modèles classiques de génétique des populations ont permis d'explorer ce lien pour des processus démographiques simplifiés. Dans cette thèse, j'examine de manière plus approfondie l'interaction entre la démographie et l'évolution génétique de populations d'agents pathogènes aux cycles de vie complexes. Pour ce faire, deux situations écologiques hors équilibre démographique sont étudiées, chacune permettant de poser des hypothèses de modélisation réalistes. L'une présente un évènement majeur de sélection où l'aspect temporel est primordial. L'autre présente des évènements récurrents de colonisation où l'aspect spatial est déterminant.

Dans une première partie, je m'intéresse à l'étude d'un contournement de résistance par une population d'agents pathogènes. En premier lieu, une approche par modélisation permet d'établir les conditions d'observation d'un contournement et d'identifier les déterminants de la durabilité d'une résistance. Les résultats mettent en évidence l'effet antagoniste de la proportion d'hôtes résistants déployés dans le paysage agricole, qui diminue la probabilité de contournement mais augmente la vitesse du contournement lorsqu'il a lieu. Ce modèle est ensuite complété par la prise en compte de l'évolution génétique aux loci neutres. Les résultats permettent d'identifier trois scénarios démographiques associés à des signatures génétiques distinctes lors du contournement : 1) De faibles variations de tailles de populations et peu d'évolution de la structure génétique, 2) Un événement fondateur fort sur l'hôte résistant qui crée en retour une structure sur l'hôte sensible et 3) Un événement de sauvetage évolutif qui se traduit par un fort événement de fondation sur l'hôte résistant, précédé par un goulot d'étranglement sur l'hôte sensible. Enfin, ce cadre théorique d'analyse démogénétique est appliqué à des données empiriques pour inférer les paramètres régissant le contournement des peupliers résistants RMlp7 par l'agent pathogène *Melampsora larici-populina*. Deux paramètres sont particulièrement bien estimés et en accord avec nos connaissances biologiques : une forte proportion d'hôtes résistants dans le paysage (plus de 80 %) et une fréquence initiale d'allèles virulents dans la population d'agents pathogènes entre 5 et 10 %.

Dans une seconde partie, je m'intéresse à l'étude de la colonisation et ses conséquences génétiques. Ces analyses portent sur un système écologique particulier, l'invasion récurrente de la vallée de la Durance par *Melampsora larici-populina*. Premièrement, un modèle mécanico-statistique est couplé aux données de suivi épidémiologique afin d'inférer les paramètres régissant la dynamique d'expansion de l'agent pathogène. Cette approche montre que la colonisation résulte de fréquents évènements de dispersion à longue distance, pour une distance moyenne de dispersion de plus de 2 km. Deuxièmement, la caractérisation de plusieurs évènements annuels de colonisation montre une structure génétique similaire mais qui varie fortement en amplitude d'une année sur l'autre. Deux extrêmes sont identifiés : en 2011, une forte conservation de la diversité génétique initiale le long du domaine colonisé ; en 2004, une érosion rapide de la diversité génétique. Les variations inter-annuelles de ces structures peuvent être expliquées par des variations dans la fréquence des évènements de dispersion à longue distance.

Ce travail permet de discuter l'importance de modèles contextualisés et prenant en compte les variations démogénétiques dans la compréhension des systèmes biologiques. Les informations obtenues sont déterminantes pour le renforcement des stratégies de lutte contre les agents pathogènes responsables de maladies émergentes.

AgroParisTech
22 place de l'Agronomie
CS 20040
F-91123 PALAISEAU cedex