



HAL
open science

Effet de la dérive génétique et de la sélection sur la durabilité de la résistance des plantes aux virus

Elsa Rousseau

► **To cite this version:**

Elsa Rousseau. Effet de la dérive génétique et de la sélection sur la durabilité de la résistance des plantes aux virus. Sciences agricoles. Université Nice Sophia Antipolis, 2016. Français. NNT : 2016NICE4024 . tel-02796455v2

HAL Id: tel-02796455

<https://theses.hal.science/tel-02796455v2>

Submitted on 28 May 2017

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.

UNIVERSITE NICE SOPHIA ANTIPOLIS

ECOLE DOCTORALE STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

THESE

pour l'obtention du grade de

Docteur en Sciences

de l'Université de Nice-Sophia Antipolis

Mention : Automatique, Traitement du Signal et des Images

présentée et soutenue par

Elsa ROUSSEAU

EFFET DE LA DÉRIVE GÉNÉTIQUE ET DE LA SÉLECTION SUR LA DURABILITÉ DE LA RÉSISTANCE DES PLANTES AUX VIRUS

Thèse dirigée par *Jean-Luc GOUZÉ* (Inria Sophia Antipolis) et par *Benoît MOURY* (INRA Avignon),
et co-encadrée par *Frédéric FABRE* (INRA Bordeaux), *Frédéric GROGNARD* (Inria Sophia Antipolis)
et *Ludovic MAILLERET* (INRA/Inria Sophia Antipolis)

soutenue le *27 mai 2016*

Jury :

M. Fernando GARCÍA-ARENAL	Professeur, CBGP Madrid	Rapporteur
M. Frédéric HOSPITAL	Directeur de Recherche, INRA Jouy-en-Josas	Rapporteur
M. Christian LANNOU	Directeur de Recherche, INRA Jouy-en-Josas	Examineur
M. Gaël THÉBAUD	Chargé de Recherche, INRA Montpellier	Examineur
Mme. Elisabeta VERGU	Chargée de Recherche, INRA Jouy-en-Josas	Examinatrice
M. Alain VIARI	Directeur de Recherche, Inria Grenoble	Examineur
M. Jean-Luc GOUZÉ	Directeur de Recherche, Inria Sophia Antipolis	Directeur
M. Frédéric FABRE	Chargé de Recherche, INRA Bordeaux	Co-encadrant

"Il faut choisir, se reposer ou être libre."

Thucydide



Remerciements

J'ai effectué cette thèse au sein de l'Unité de Pathologie Végétale à l'INRA d'Avignon pour ma première année, puis dans l'équipe Biocore d'Inria Sophia Antipolis pour les deux années et demie suivantes. Cette thèse a été financée grâce à une bourse CORDI Inria.

Je remercie l'équipe de direction de l'Unité de Pathologie Végétale, Marc Bardin, Cindy Morris, Christel Leyronas et Benoît Moury, ainsi que celle de l'équipe Biocore, Jean-Luc Gouzé et Olivier Bernard, pour m'avoir accueillie et permis de réaliser ma thèse dans de si bonnes conditions.

Je remercie tous les membres de mon comité de thèse, Samuel Alizon, Stéphane Blanc, Jérôme Enjalbert, Étienne Klein et Karine Berthier pour leur implication, pour des échanges très constructifs et beaucoup de bons conseils.

Je tiens sincèrement à remercier Fernando García-Arenal et Frédéric Hospital pour avoir accepté d'être rapporteurs de ma thèse, et merci pour vos excellents rapports ! Merci à Christian Lannou, Gaël Thébaud, Elisabeta Vergu et Alain Viari d'avoir accepté de faire partie de mon jury de thèse. Je suis très honorée d'avoir un tel jury pour ma soutenance de thèse.

Je remercie infiniment mes quatre encadrants de thèse, Frédéric Fabre (INRA Bordeaux), Frédéric Grognaud (Inria Sophia Antipolis), Ludovic Mailleret (INRA et Inria Sophia Antipolis) et Benoît Moury (INRA Avignon, co-directeur de thèse), pour m'avoir permis de réaliser cette thèse avec eux. Merci à Frédéric F pour ton dévouement, tout le temps que tu as pris pour moi, que ce soit pour des longues heures de discussion lorsque j'avais des doutes ou des questions, pour ton aide pour les simulations ou encore lorsqu'il fallait mettre la main à la pâte pour les manips. Merci à Frédéric G pour ton dévouement également, ce sont nos deux bureaux qui étaient les plus proches pendant la majeure partie de ma thèse, c'est donc toi que j'allais souvent voir en premier quand j'avais une question ou besoin d'aide ! Je te remercie pour ta disponibilité et tous tes bons conseils. Petit clin d'œil à la fin de ma thèse : à quand ton HDR ?;-) Merci à Ludovic pour ton investissement et toutes tes remarques et conseils toujours pertinents. Tous les moments que nous avons échangés ont toujours été très constructifs pour moi et m'ont réellement aidé à avancer. Merci à Benoît pour ta culture scientifique et générale démesurée, c'est un réel plaisir de discuter avec toi ! Tu m'as énormément appris, notamment sur la biologie des virus et des plantes, j'ai adoré découvrir ces domaines à tes côtés ! Enfin, merci à vous quatre pour votre investissement et votre soutien, qui m'ont permis de beaucoup avancer, et de mener à bien cette thèse. Certes, l'approche de la fin de thèse a parfois quelque peu "pimenté" nos réunions, mais nous avons finalement toujours réussi à trouver une entente, et à aller de l'avant. Au final, je garderai une image très positive de cette thèse, et j'espère retrouver des collègues comme vous pour la suite !

Je remercie également mes autres encadrants (oui, j'en ai eu beaucoup !), Alain Palloix, Rachid

Senoussi et Jérôme Coville, pour tous les échanges constructifs qui m'ont beaucoup apporté, et pour vous être impliqués dans ma thèse. Merci Rachid pour tout le temps que tu as pris pour m'aider sur des problèmes de statistiques, et merci à Rachid et Jérôme pour votre participation aux réunions et comités de thèse. Mes pensées vont tout particulièrement à Alain, dont le décès survenu le 8 mars 2016 nous a tous beaucoup affecté. J'ai sincèrement apprécié ta présence au cours de ma thèse et à nos réunions, auxquelles tu amenais toujours des discussions passionnantes qui m'ont beaucoup enrichie.

Un grand merci à l'équipe de Virologie. Je remercie tout particulièrement Vincent et Benoît pour m'avoir formée aux manips de biologie moléculaire. Merci à Vincent, Baptiste et Greg pour votre aide précieuse pour les manips. Merci à l'équipe des serres, Nathalie, Michèle, Joël, Laurent et Michel pour tout le soin que vous apportez aux plantes. Un clin d'œil à Nathalie et Michèle qui m'ont particulièrement aidée lors des grosses manips, toujours dans la bonne humeur! Merci à toute l'équipe de Pathologie Végétale pour des pauses café toujours conviviales et presque toujours accompagnées de gateaux faits maison délicieux. Merci pour votre accueil chaleureux!

Je remercie l'équipe Résistance Durable chez les Solanacées, qui a fourni tout le matériel 'graines' pour réaliser mes expériences. Je remercie tout particulièrement Ghislaine pour son aide et sa bienveillance.

Je tiens à remercier toute l'équipe Biocore pour son accueil. Particulièrement, merci à Suzanne Touzeau pour ta disponibilité et ta bienveillance. Tu étais toujours disponible pour m'aider et pour discuter, pour des questions d'informatique ou en lien avec ma thèse. Merci à Jean-Luc, Olivier, Madalena, Valentina et Pierre.

Je remercie les personnes s'occupant de l'administration à l'INRA comme chez Inria, Pascale, Claudine et Stéphanie, pour leur aide précieuse et leur bonne humeur.

À Avignon, je remercie tout particulièrement Julie, ancienne doctorante de Benoît et dont mon sujet de thèse fait suite à ses travaux. Merci pour ton dévouement pour la "transmission" de tes travaux. Je remercie également Sébastien, stagiaire de Master 2 que j'ai participé à encadrer, pour son investissement et sa disponibilité. On a réalisé des manips très "lourdes" ensembles, mais tu étais toujours volontaire pour rester plus longtemps pour m'aider à terminer, toujours dans la bonne humeur. Merci à Baptiste, Karine, Jonathan, Hana, Yuko, Julie et Sébastien pour tous ces bons moments passés ensemble et tous ces bons repas partagés! Enfin, merci à toute l'équipe de volley de l'INRA d'Avignon, pour des bonnes soirées d'entraînement et pour le super barbeuc de fin d'année!

À Sophia Antipolis, je remercie tous les doctorants, post-doctorants et ingénieurs des équipes Biocore et Apics : Mélaïne, Mélanie, Nicolas, David, Claudia, Stefano, Francesco, Carlos, Ignacio, Quentin, Francis, Alfonso, Camille, Pierre-Olivier, Sofia, Christos, Konstantinos, Dmitry, Bapan, Florentina. Merci pour les super pauses café passées avec vous, toujours remplies de bonne humeur et de rires, ingrédients nécessaires pour réussir sa thèse! Un clin d'oeil particulier à toute l'"équipe"

escalade, pour des sessions de grimpe sportives, mais aussi pleines de rires et de bons moments. Merci à Mélaine pour m'avoir fait découvrir la salle d'escalade, et pour tes super cours de Tai-Chi et de Kung-Fu! Merci à Camille pour avoir toujours été à l'écoute, et pour m'avoir donné des bons conseils pour aller de l'avant. J'ai eu de la chance d'avoir un super "co-thésard", Nicolas, merci pour tous tes conseils, tous les bons moments et les fous-rires. Je remercie ma stagiaire de Master 1, Mélanie Bonneault, pour son travail très appréciable pour la conception et les premières simulations du modèle épidémiologique, c'était un réel plaisir de travailler avec toi. Je te souhaite toute la réussite possible pour la fin de tes études et le début de ta vie professionnelle. Merci aussi à ma prof de yoga, qui m'a initiée à cette activité en dernière année de thèse, et je compte bien continuer!

Merci au virus Y de la pomme de terre (PVY) et au piment, pour m'avoir permis de réaliser toutes mes expériences et de trouver des résultats très intéressants pour ma thèse! Et merci au PVY fluorescent pour m'en avoir fait voir de toutes les couleurs!

Je remercie mes amis, Raphaëlle pour tous les bons moments, pour avoir toujours été là quand j'en avais besoin, et pour m'avoir appris à tricoter, très bonne activité de relaxation en fin de thèse! Merci à Audrey, même si tu étais à l'autre bout du monde, tous nos échanges par mail ou par skype, et tes visites en France m'ont toujours réchauffé le cœur. Merci à tous les BIM (INSA Lyon) pour les week-end que l'on continue à organiser pour se retrouver et qui me font toujours le plus grand bien : Alice, Greg, Hugo, Béré, Sacha, Elsa, Sam, Jouis, Romain, Léa, Raphaëlle, Audrey...

Enfin, un grand merci à ma famille : mes parents Françoise et Xavier, mon frère Maxime et sa femme Audrey et leur petite Raphaëlle, née pendant ma thèse, et qui m'a apporté beaucoup de bonheur et de joie. Merci à mon copain Auré pour tous les bons moments passés à tes côtés, pour ton attention et tes bons petits plats. Merci à mes cousins, oncles et tantes : Nico, Sarah, Léa, Chris, Louis, Charlotte, Clément, Caro, Robert, Jean-Pierre, Patrice, Zouzou, Pierre, Claire, pour toutes les super cousinades et moments partagés. Un merci particulier à Léa, Sarah et Nico pour votre disponibilité et votre attention.

J'ai certainement oublié des personnes dans cette liste. Je m'excuse donc par avance pour ces oublis, et je remercie pour terminer toutes les personnes que j'ai côtoyées pendant ma thèse et avec qui j'ai partagé des bons moments.

Liste des abréviations

- AUDPC** aire sous la courbe de progression de la maladie (*Area Under the Disease Progress Curve*). 32
- BYDV** *Barley yellow dwarf virus* (genre *Luteovirus*). 194
- CaMV** *Cauliflower mosaic virus* (genre *Caulimovirus*). 39, 40
- CMV** *Cucumber mosaic virus* (genre *Cucumovirus*). 40, 58, 186
- CTMC** chaînes de Markov en temps continu (*Continuous Time Markov Chain*). 63–66
- DTMC** chaînes de Markov en temps discret (*Discrete Time Markov Chain*). 63
- HD** haploïde doublée. 31–34, 174–176, 186
- HIV** *Human immunodeficiency virus* (genre *Lentivirus*). 75
- IBM** modèle individu-centré (*Individual Based Model*). 63
- MNSV** *Melon necrotic spot virus* (genre *Carmovirus*). 38, 39
- ODE** équations différentielles ordinaires (*Ordinary Differential Equations*). 61–65
- PVX** *Potato virus X* (genre *Potyvirus*). 38
- PVY** *Potato virus Y* (genre *Potyvirus*). 3, 12, 18, 21, 29–34, 36–40, 50, 186, 187
- QTL** loci à effet quantitatif (*Quantitative Trait Loci*). 28–33, 186, 187, 190, 191, 194
- SDE** équations différentielles stochastiques (*Stochastic Differential Equations*). 63
- SIDA** syndrome d'immunodéficience acquise. 75
- TEV** *Tobacco etch virus* (genre *Potyvirus*). 39, 40
- TMV** *Tobacco mosaic virus* (genre *Tobamovirus*). 18, 40

ToMV *Tomato mosaic virus* (genre *Tobamovirus*). 38

TSWV *Tomato spotted wilt virus* (genre *Tospovirus*). 16, 18, 38, 58

TuMV *Turnip mosaic virus* (genre *Potyvirus*). 38, 39

WSMV *Wheat streak mosaic virus* (genre *Tritimovirus*). 40

Table des matières

Remerciements	5
Liste des abréviations	7
1 Résistances des plantes et forces évolutives	17
1.1 Contexte et enjeux agro-économique et agro-écologique	17
1.2 Résistance qualitative des plantes aux agents pathogènes	20
1.2.1 Le contournement d'une résistance depuis l'échelle de la cellule jusqu'à celle du paysage	22
1.2.2 Durabilité d'une résistance	27
1.2.3 Monocultures et risques d'épidémies	31
1.2.4 Un enjeu majeur : intégrer les principes éco-évolutifs régissant les interactions hôtes - agents pathogènes pour une meilleure gestion des cultures.	34
1.3 Résistance quantitative des plantes aux agents pathogènes	38
1.3.1 Le fond génétique de la plante peut moduler le taux de contournement d'un gène majeur de résistance	40
1.3.2 Trois mécanismes pour expliquer la plus grande durabilité de la résistance polygénique	43
1.4 Forces évolutives et adaptation à une résistance	47
1.4.1 Quelques caractéristiques des agents pathogènes en lien avec leur adaptation à résistance des plantes	48
1.4.2 Interaction entre dérive génétique et sélection	52
1.4.3 Hypothèses sur le lien entre la fréquence de contournement du gène majeur de résistance <i>pvr2³</i> du piment et les forces évolutives agissant sur le <i>Potato virus Y</i> (genre <i>Potyvirus</i>) (PVY)	61
1.4.4 Applications à l'échelle d'un agro-écosystème	62

2	Épidémiologie, modélisation et agriculture	65
2.1	Définitions	65
2.2	Épidémiologie et caractéristiques des parasites	66
2.2.1	Épidémiologie animale et humaine	66
2.2.2	Épidémiologie végétale	68
2.2.3	Micro- et macro-parasites	69
2.2.4	Modes de transmission	71
2.2.5	Définition du cadre d'étude	71
2.3	Introduction aux modèles à compartiments	72
2.3.1	Le modèle déterministe SIR de Kermack et McKendrick	72
2.3.2	Version stochastique d'un modèle SIS simple	75
2.4	Durabilité de la résistance des plantes	79
2.4.1	Caractéristiques liées à l'épidémiologie dans les agro-écosys-tèmes	79
2.4.2	Stratégies de gestion du paysage	83
2.4.3	Épidémiologie évolutive	86
	Structure de la thèse	91
3	Genetic drift at the host population scale	93
3.1	Introduction	95
3.2	Materials and methods	98
3.2.1	Model overview	98
3.2.2	Description of the deterministic model	99
3.2.3	Stochastic model description	101
3.2.4	Model analysis	103
3.3	Results	105
3.3.1	Epidemic dynamics	105
3.3.2	Impact of the choice of the resistance gene	107
3.3.3	Global impact of all the factors	110
3.4	Discussion	112
3.4.1	Global impact of the choice of the resistance gene on the additional relative benefit	113
3.4.2	Impact of the four factors	114

3.4.3	Model-imposed link between the proportion of resistant plants and the intensity of genetic drift	116
3.4.4	Future developments of the model	118
3.4.5	Extending the spatial and temporal scales of the model	119
3.4.6	Concluding remarks	120
4	Joint Estimation of Genetic Drift and Selection	121
4.1	Introduction	125
4.2	Results	127
4.2.1	Overview of the method	127
4.2.2	Numeric experiment with FFPopSim	131
4.2.3	Range of selection and genetic drift intensities explored in numerical simulations	133
4.2.4	Estimation of effective population size and variant fitness from simulated datasets	134
4.2.5	Observed dynamics of variant frequencies in real-life experiments with plant virus populations	137
4.2.6	Estimation of effective population sizes and variant fitness from real-life datasets	139
4.3	Discussion	141
4.3.1	A method for the joint estimation of genetic drift and selection over a large range of intensities from experimental evolution data	141
4.3.2	Plant genotypes modulate genetic drift and selection within virus populations.	145
4.4	Materials and Methods	147
4.4.1	Biological experiment	147
4.4.2	Maximum-likelihood estimation	148
5	Factors influencing adaptation to plant resistance	163
5.1	Introduction	165
5.2	Results	168
5.2.1	Estimation of the frequency of breakdown of the major resistance gene <i>pvr2³</i> by PVY and of putative explanatory variables linked to PVY evolution	168
5.2.2	Correlation between putative explanatory variables	173

5.2.3	Analysis of the links between the frequency of breakdown of the major resistance gene and putative explanatory variables	176
5.3	Discussion	177
5.3.1	Genetic drift and resistance efficiency are complementary levers to improve the durability of virus resistance genes	177
5.3.2	Complex interplays between virus within-host accumulation, genetic drift and selection	178
5.3.3	Concluding remarks	181
5.4	Materials and methods	182
5.4.1	Analysis of composite PVY populations infecting pepper DH lines	182
5.4.2	Quantitative estimation of PVY accumulation in pepper plants	183
5.4.3	Inference of virus effective population size and selection coefficients	184
5.4.4	Statistical analyses of the links between variables related to the evolution of PVY populations	184
6	Synthèse, discussion et perspectives	187
6.1	Synthèse des principaux résultats	187
6.2	Discussion	190
6.2.1	Intensité de la dérive génétique et taille de la population d'hôtes	190
6.2.2	Mutation et probabilité d'émergence de mutants adaptés	192
6.2.3	Pertinence du modèle d'estimation de N_e et s ?	197
6.2.4	Taille efficace de population et taux d'évolution d'une population : limites actuelles	198
6.2.5	Variabilité phénotypique due au génome de l'agent pathogène	199
6.2.6	Application à d'autres agents pathogènes?	201
6.3	Perspectives	202
6.3.1	Identifier des QTL de taille efficace de populations virales chez les plantes hôtes	202
6.3.2	Modèle d'estimation de N_e et s	203
6.3.3	Améliorations du modèle épidémiologique	203
6.4	Remarques finales	207
	Bibliographie	212
	Liste des publications, présentations et posters	243

Chapitre 1

Résistances des plantes et forces évolutives

1.1 Contexte et enjeux agro-économique et agro-écologique

Face à une population humaine toujours en croissance, la *sécurité alimentaire*¹ est un défi que nous devons relever tout en veillant à limiter l'impact des pratiques agricoles sur l'environnement (Bommarco et al., 2013; Tscharntke et al., 2012). La taille de la population humaine devrait augmenter jusqu'à atteindre un plateau à environ 9 milliards d'individus aux alentours de 2050, période vers laquelle le monde aurait besoin de 70 à 100% de nourriture supplémentaire par rapport à aujourd'hui (Food and Agriculture Organisation (FAO) of the United Nations, 2009; Godfray et al., 2010). La production agricole a progressé de manière régulière et linéaire au cours des cinquante dernières années, mais le taux de progression doit augmenter de 38% durant les quarantes années à venir pour assurer l'alimentation humaine (Tester and Langridge, 2010). Le défi à relever comporte trois facettes : (i) veiller à répondre à une demande en nourriture toujours plus forte face à la croissance de la population humaine, (ii) minimiser l'impact environnemental et social d'une telle demande, et (iii) assurer que les personnes les plus pauvres ne souffrent plus de malnutrition, ce qui concerne environ 2 milliards de personnes dans le monde aujourd'hui (Godfray et al., 2010; Tscharntke et al., 2012).

1. Le Comité de la Sécurité Alimentaire mondiale a défini la sécurité alimentaire comme un accès physique et économique, à tout moment et pour tout être humain, à une nourriture suffisante, saine et nutritive leur permettant de satisfaire leurs besoins nutritionnels et leurs préférences alimentaires pour une vie active et saine (Food and Agriculture Organisation (FAO) of the United Nations, 1996).

La diminution du gaspillage des aliments est un point crucial pour la sécurité alimentaire. En effet, environ 30 à 40% de la nourriture produite est actuellement perdue. Dans les pays en voie de développement, le gaspillage est principalement dû à des problèmes de stockage dans les fermes ainsi que pendant le transport. Dans les pays développés, la part principale des pertes est observée depuis la commercialisation jusqu'à la consommation (Godfray et al., 2010; Gustavsson et al., 2011; Movahedi et al., 2009; Tscharntke et al., 2012).

La diminution de l'impact environnemental dû aux pratiques agricoles est également un point clé. À l'aube de l'an 2000, Tilman rapportait un doublement de la production agricole de nourriture au cours des 35 années précédentes (Tilman, 1999). Malheureusement, cette augmentation était associée à l'utilisation d'environ 7 fois plus d'engrais azotés, 3 à 4 fois plus d'engrais phosphorés et 1 à 2 fois plus de terres irriguées pour quasiment la même surface de terres cultivées (Tilman, 1999). La réduction de l'utilisation de l'eau est un des enjeux majeurs pour demain, notamment dans les régions qui présenteront une plus grande sécheresse à cause du dérèglement climatique (Godfray et al., 2010; Tester and Langridge, 2010). L'utilisation massive d'engrais chimiques, mais également de produits phytosanitaires, a conduit par ruissellement à des impacts négatifs très importants sur les écosystèmes d'eaux douces (Moss, 2008), et a rendu de nombreuses eaux impropres à la consommation (Tilman, 1999). En ce qui concerne les produits phytosanitaires, leur effet néfaste global sur l'environnement et sur la biodiversité n'est plus à démontrer (Bommarco et al., 2013; Potts et al., 2010). La législation limite aujourd'hui de plus en plus l'utilisation des engrais chimiques pour l'agriculture (Tester and Langridge, 2010). En Europe, la directive pesticide de 2009 régit leur usage et en France, le plan Ecophyto II (faisant suite au plan Ecophyto 2018) a été mis en place en réponse au Grenelle de l'Environnement, visant à réduire de moitié l'utilisation des pesticides entre 2008 et 2025.

L'augmentation des surfaces cultivées pour l'agriculture ne paraît pas être une bonne solution face à la compétition pour la terre venant des autres activités humaines, ainsi que face à la nécessité de protéger les écosystèmes naturels pour leur biodiversité ainsi que pour leur capacité à fixer le carbone (Godfray et al., 2010). Les activités humaines en compétition pour la terre concernent notamment la production de bio-carburants et l'urbanisation (Movahedi et al., 2009; Tscharntke et al., 2012). À titre d'exemple, d'ici à 2020, 15% des productions de céréales et d'huiles végétales ainsi que 30% de la production de canne à sucre devraient servir pour les bio-carburants (Tscharntke et al., 2012). À l'opposé, les surfaces de terre cultivées devraient augmenter de seulement 8% jusqu'en 2030, après quoi nous arriverons à saturation,

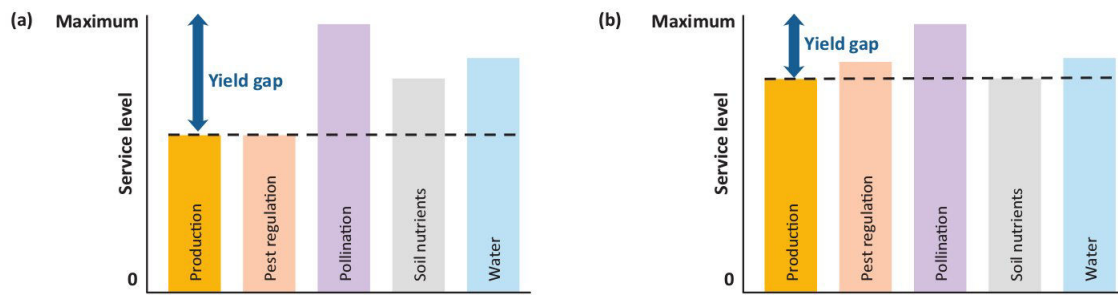


FIGURE 1.1 – Illustration de la contribution d’une régulation des agents pathogènes à la production agricole et à la réduction de l’écart de rendement (Yield gap). (a) La production agricole (Production) ne peut pas aller au-delà d’un niveau fixé par le service de régulation sous-jacent le plus faible, ici en l’occurrence la régulation des ravageurs et agents pathogènes (Pest regulation). (b) La régulation de ces derniers étant améliorée, la production agricole augmente et l’écart de rendement est réduit jusqu’à un seuil fixé par le service de régulation le plus faible suivant, dans cet exemple les nutriments présents dans le sol. Source : (Bommarco et al., 2013).

et les surfaces ajoutées d’ici là devraient tout juste compenser les terres abandonnées et celles affectées à l’urbanisation (Movahedi et al., 2009). Ainsi, une augmentation de la production agricole devra passer par une augmentation du rendement des cultures (Movahedi et al., 2009).

Une manière de procéder pour augmenter les rendements sur une même surface agricole consiste à tenter de diminuer l’écart de rendement. Cette notion d’écart de rendement fait référence à la différence entre la productivité réalisée et la meilleure productivité pouvant être atteinte en utilisant les ressources génétiques et les méthodes de gestion à disposition (Godfray et al., 2010). La meilleure productivité pouvant être atteinte à un endroit donné dépend de l’utilisation de différents services d’appui et de protection, tels que la régulation des bioagresseurs, la pollinisation, les nutriments présents dans le sol ou l’eau (Bommarco et al., 2013). En jouant sur ces services d’appui et de protection, la productivité pourrait être améliorée et donc l’écart de rendement réduit (Fig. 1.1).

Plus généralement, l’ensemble des solutions visant à produire davantage sur la même surface agricole tout en réduisant l’impact environnemental a été nommé *intensification durable* (Garnett et al., 2013; Godfray et al., 2010). Cette notion inclut, entre autres, la recherche de variétés de plantes plus productives et adaptées à un apport mineur en eau, ou la mise en place de stratégies de protection des cultures intégrées contre les ravageurs et agents pathogènes (Movahedi et al., 2009). Globalement, parmi l’ensemble des agents pathogènes des plantes (virus, champignons, bactéries, nématodes, etc.), les virus sont responsables de 47% des maladies infectieuses émergentes, les champignons de 30% et les bactéries de 16% (Anderson et al., 2004).

Le changement climatique global et rapide ajoute une difficulté supplémentaire au défi

d'accroissement de la production agricole (Garrett et al., 2011; Tester and Langridge, 2010). En effet, le changement climatique se traduisant notamment par une hausse globale de la température, une irrégularité du climat, ainsi qu'une diminution de la disponibilité en eau (European Environment Agency (EEA), 2012), demandera des variétés de plantes adaptées à de telles conditions, mais devrait également avoir un impact sur les populations d'agents pathogènes ravageant les cultures, qu'il faudra anticiper (Garrett et al., 2011).

Le sujet de cette thèse se situe au cœur des deux concepts introduits précédemment, à savoir l'écart de rendement et l'intensification durable. Elle s'inscrit dans une optique d'augmentation de la productivité en jouant sur la régulation des agents pathogènes. Sans faire appel à la protection chimique avec des produits phytosanitaires, cette régulation se base sur le contrôle génétique via l'utilisation de variétés de plantes résistantes aux agents pathogènes. Cette thèse s'intéresse plus particulièrement aux principes éco-évolutifs régissant les interactions entre hôtes et agents pathogènes pour un meilleur déploiement des variétés résistantes dans le paysage agricole. En effet, tous les organismes sont en constante évolution et l'on peut s'attendre à ce que les agents pathogènes s'adaptent tôt ou tard à la résistance des plantes (Parlevliet, 2002). Jouer sur ces principes évolutifs pourrait permettre de retarder cette adaptation, et donc de maintenir une meilleure productivité sur le long terme. Dans la suite du document, nous verrons d'abord ce qu'est une résistance qualitative et comment elle est contournée, puis nous verrons un deuxième type de résistance qui peut être bénéfique à une meilleure gestion des populations d'agents pathogènes et pour augmenter la productivité. Enfin, nous verrons par quels moyens jouer sur les forces gouvernant l'évolution des agents pathogènes afin d'en améliorer la gestion.

1.2 La résistance des plantes comme méthode de protection contre les agents pathogènes

Les résistances génétiques des plantes aux agents pathogènes existent depuis toujours dans les plantes sauvages. Dès les années 1930, le contrôle génétique pour la protection contre les bioagresseurs s'est développé, reposant sur la création de variétés porteuses de gènes de résistance (Holmes, 1938; Moury et al., 2010). Ces résistances *monogéniques*, reposant sur un gène de résistance majeur vis-à-vis d'un agent pathogène, sont également appelées résistances *qualitatives* ou totales. Leur mécanisme d'action au niveau populationnel peut être résumé

par le modèle gène-pour-gène ou par le modèle matching allele (Flor, 1971; Moury et al., 2010; Sacristán and García-Arenal, 2008). Dans les versions simple-locus biallélique de ces deux modèles (Table 1.1), la plante possède un gène de résistance avec deux allèles, un allèle de résistance (R) et un allèle de sensibilité (S), et l'agent pathogène possède un gène d'avirulence avec deux allèles, un allèle d'avirulence (avr) et un allèle de virulence (vir). L'interaction entre la protéine produite par le gène de résistance d'une part et celle produite par le gène d'avirulence d'autre part détermine la réussite ou non de l'infection. Dans le modèle simple-locus décrit, quatre interactions sont possibles entre la plante et son agent pathogène. Pour le modèle gène-pour-gène, la seule interaction conduisant à une incompatibilité, c'est-à-dire à un blocage de l'agent pathogène par la plante, correspond au cas où un agent pathogène avr tente d'infecter une plante R . Les trois autres scénarios d'interactions conduisant à l'infection des plantes, ces interactions sont dites compatibles (Brown and Tellier, 2011). Pour le modèle matching allele, seule la diagonale de la Table 1.1 conduit à des interactions compatibles. Cela signifie que l'acquisition de l'allèle de virulence par l'agent pathogène entraîne une perte de compatibilité entre celui-ci et la plante sensible.

		Agent pathogène	
		avirulent	virulent
Plante	Sensible	1	GPG : 1/MA : 0
	Résistante	0	1

TABLE 1.1 – **Tableau présentant les modèles gène-pour-gène et matching allele.** L'agent pathogène a deux variants : avirulent ou virulent. La plante hôte possède deux variétés : sensible (dépourvue du gène majeur de résistance) ou résistante. Les interactions entre l'agent pathogène et sa plante hôte conduisant à l'infection de la plante sont représentées par un 1, celles ne conduisant pas à l'infection par un 0. Dans le modèle gène-pour-gène (GPG), le variant virulent est capable d'infecter les plantes sensibles alors que dans le modèle matching allele (MA) ce n'est pas le cas.

Au niveau moléculaire, dans le modèle gène-pour-gène c'est la reconnaissance entre la protéine produite par l'allèle de résistance chez la plante et la protéine produite par l'allèle d'avirulence chez le virus qui déclenche une réaction de défense chez la plante, conduisant à une limitation de la multiplication de l'agent pathogène depuis son site d'infection, souvent par mort cellulaire liée à une *réaction d'hypersensibilité*². À l'inverse, dans le modèle matching allele seule une parfaite reconnaissance entre les deux protéines produites par la plante d'une part et par l'agent pathogène d'autre part conduit à l'infection. Dans le modèle matching allele, les agents pathogènes sont donc des spécialistes de leur hôte, alors que dans le modèle gène-pour-gène l'acquisition d'allèles de virulence permet à l'agent pathogène de devenir de plus en

2. Mort cellulaire rapide et localisée au site d'infection dans les interactions de résistance entre une plante et un agent pathogène (D'Arcy et al., 2001)

plus généraliste. Même si historiquement le modèle gène-pour-gène était attribué aux systèmes végétaux et le modèle matching allele aux systèmes animaux, on sait aujourd'hui que les deux types de modèles sont pertinents pour expliquer les interactions plantes - agents pathogènes (Sacristán and García-Arenal, 2008). Par exemple, l'interaction piment - tobamovirus est bien décrite par le modèle gène-pour-gène alors que l'interaction entre le gène de résistance *pvr2* chez le piment et le PVY, étudiée sur huit allèles côté virus et côté plante, ne correspond ni complètement au modèle matching allele ni complètement au modèle gène-pour-gène (Moury et al., 2014). Dans la suite du document, par souci de simplicité je ne parlerai que du modèle gène-pour-gène (Flor, 1971).

Une précision s'impose à ce stade concernant le terme de *virulence*, car il n'a pas le même sens en phytopathologie et en biologie évolutive. En phytopathologie, la virulence est définie comme la capacité de l'agent pathogène à infecter une plante (composante binaire, comme pour le modèle gène-pour-gène), alors qu'en biologie évolutive elle correspond à l'effet néfaste sur l'hôte de l'infection par l'agent pathogène (trait quantitatif, par exemple la sévérité des symptômes) (Lannou, 2012; Moury et al., 2010). Ici, j'emploierai le terme de virulence dans son sens qualitatif, et je parlerai d'*agressivité* pour le sens quantitatif. Plus généralement, la pathogénicité est un terme à l'interface entre virulence et agressivité, et est définie comme l'aptitude à causer une maladie (Lannou, 2012).

1.2.1 Le contournement d'une résistance depuis l'échelle de la cellule jusqu'à celle du paysage

L'introduction d'une variété résistante dans un paysage agricole peut s'accompagner du contournement de la résistance par l'agent pathogène, c'est-à-dire de l'émergence et de la propagation d'un variant *vir*. Le contournement d'une résistance est un phénomène faisant intervenir divers processus à différentes échelles d'espace et de temps (Fig. 1.2). Je vais ici présenter ces étapes dans le cas des virus de plante, mais elles peuvent facilement être généralisées à d'autres pathosystèmes. La première étape consiste en l'apparition d'un variant *vir* par mutation ou recombinaison dans une cellule d'une plante hôte (Encadré 1.1). La deuxième étape correspond à la colonisation de la plante, dans laquelle les variants *vir* et *avr* vont être en compétition et seront soumis aux forces évolutives de sélection (Encadré 1.2) et de dérive génétique (Encadré 1.3). La sélection est liée à la fitness³ de chaque variant. Le variant qui

3. Capacité d'un individu à survivre et à se reproduire, ou de manière similaire capacité d'un individu à transmettre ses gènes à la génération suivante (D'Arcy et al., 2001).

aura la plus forte fitness sera le plus avantage, il tendra donc à augmenter en fréquence dans la population. La dérive génétique se caractérise par des fluctuations aléatoires dans l'évolution des fréquences des variants. Plus la dérive génétique est forte, plus les fluctuations sont importantes, pouvant parfois conduire à l'extinction de variants. Le variant *vir* doit donc être suffisamment compétitif et ne pas disparaître par dérive génétique pour pouvoir être disséminé à l'échelle du paysage, par migration (Encadré 1.4). Lors de cette troisième étape, les populations d'agents pathogènes sont soumises à la dérive génétique, à la sélection et à la migration. Enfin, dans les climats tempérés comme le nôtre, une quatrième étape intervient, à savoir la saisonnalité. En effet, la récolte des cultures de plantes annuelles annonce la disparition de ces plantes du paysage pendant l'hiver. Pour pouvoir persister, le variant *vir* doit trouver un moyen de survivre à l'hiver, par exemple dans les plantes sauvages ou sous une forme résistante au froid (Agrios, 2005; Mailleret et al., 2012).

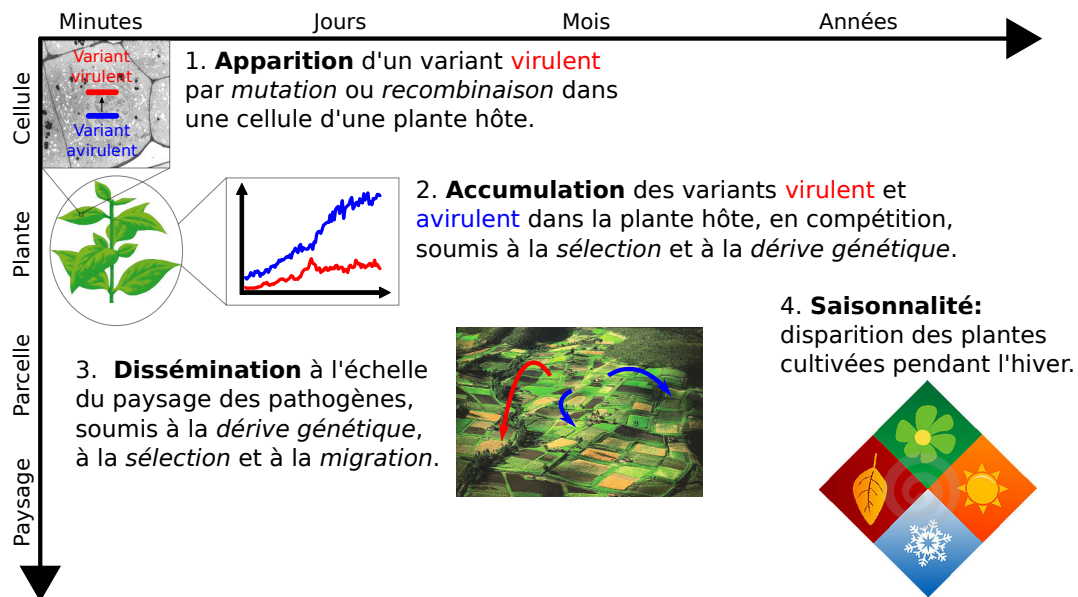


FIGURE 1.2 – Illustration des quatre étapes du contournement d'une résistance par un virus, aux différentes échelles d'espace (axe vertical) et de temps (axe horizontal) impliquées. Modifiée d'après Moury et al. (2010).

ENCADRÉ 1.1

La **mutation** est la source principale de variation génétique. Elle correspond à la modification d'un ou de plusieurs nucléotides dans le génome d'un individu. Ce processus permet de créer de nouveaux variants virulents chez les agents pathogènes qui pourront conduire au contournement d'une résistance (McDonald and Linde, 2002a). Il existe trois types de mutations : les substitutions (changement d'un nucléotide), et les insertions et délétions de nucléotides. On distingue enfin deux types de substitutions, à savoir les substitutions non-synonymes ou

synonymes, modifiant ou ne modifiant pas (respectivement) l'acide aminé codé.

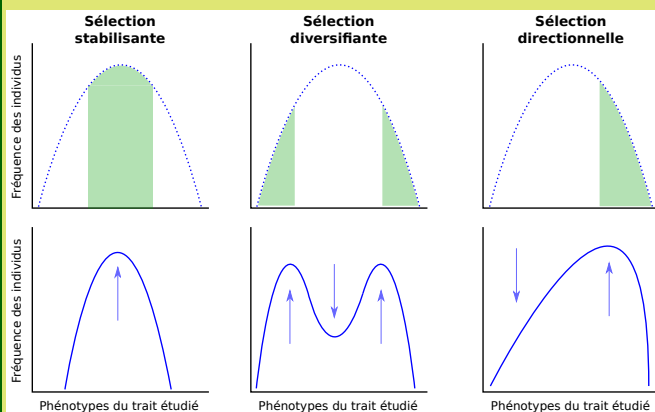
La **recombinaison** correspond à un échange de fragments de génome entre deux génomes différents, créant ainsi de nouveaux variants génétiques. Selon le type d'organisme, la recombinaison se produit de différentes façons : lors de la reproduction sexuée (pendant la méiose) chez les eucaryotes (dont font partie les champignons), lors de la conjugaison bactérienne chez les procaryotes, lors de l'anastomose ^a hyphale ^b et/ou de la recombinaison parasexuelle chez les champignons, et au sein d'une cellule infectée par plusieurs variants chez les virus (McDonald and Linde, 2002a).

a. Fusion entre les branches d'une ou de plusieurs structures (ici des hyphes) pour former un réseau (D'Arcy et al., 2001).

b. Hyphe : filament tubulaire de la thalle ou du mycélium d'un champignon; unité structurelle basique d'un champignon (D'Arcy et al., 2001).

ENCADRÉ 1.2

La **sélection** favorise systématiquement les individus caractérisés par une plus forte fitness³ d'une génération à la suivante, aux dépens de ceux ayant une plus faible fitness. Ainsi, la sélection s'exerce sur les individus, mais l'action de la sélection peut être évaluée au niveau des gènes portés par ces individus. En effet, l'intensité de la sélection varie le long du génome, et un coefficient de sélection peut être attribué à une mutation, correspondant à la différence de fitness entre un individu portant la mutation et un individu sans la mutation, toutes choses égales par ailleurs. La sélection peut être de trois types, stabilisante, diversifiante ou directionnelle. La figure ci-dessous illustre l'effet de ces trois types de sélection sur la distribution d'un trait phénotypique associé à une population.



Les graphiques sur la 1^{ère} ligne représentent la même distribution initiale d'un trait phénotypique dans trois populations localisées dans trois environnements distincts. Les zones vertes indiquent les phénotypes les plus adaptés à leur environnement, i.e. ceux qui conduisent aux meilleures fitness des individus. Les graphiques de la 2^{ème} ligne montrent l'évolution des distributions du trait phénotypique dans les trois environnements, représentatifs d'une sélection stabilisante, diversifiante ou directionnelle. Modifié d'après Quenouille (2013)

La sélection *stabilisante* tend à favoriser les phénotypes moyens par rapport aux phénotypes extrêmes. Elle réduit la variance d'un trait sans changer sa moyenne. Un exemple assez récent concerne la masse corporelle chez les oiseaux sociaux, *Philetairus socius*. Selon Covas et al. (2002), la masse intermédiaire de ces oiseaux aurait pour origine un compromis entre le risque de famine pour des individus à faible masse corporelle (moins de réserves énergétiques) et le risque de prédation pour des individus à forte masse corporelle (moins d'agilité pour éviter les

prédateurs). À l'inverse, la sélection *diversifiante*, ou disruptive, favorise deux valeurs extrêmes du trait phénotypique et élimine les intermédiaires. Par exemple, une population de prédateurs avec deux proies différentes peut conduire à la sélection de deux types de prédateurs, chacun étant spécialisé sur une proie. Les prédateurs présentant un phénotype intermédiaire sont éliminés car moins habiles sur chacune des proies. Enfin, la sélection *directionnelle* favorise des mutations qui vont toutes dans le même sens pour un trait phénotypique, elle agit sur la moyenne de la distribution d'un trait. L'exemple le plus connu est celui du 'mélanisme industriel' en Angleterre, impliquant la phalène du bouleau (*Biston betularia*) (Kettlewell, 1955). Dans une région d'Angleterre, deux formes de phalène du bouleau co-existaient, les blanches et les noires, les formes blanches étant majoritaires. Au bout d'un moment, la tendance s'inversa et les formes noires devinrent majoritaires. Kettlewell expliqua que les émissions de fumée dues à l'industrialisation naissante entraînaient un changement de la couleur des bouleaux, qui devinrent plus foncés. Les insectes clairs étaient alors plus visibles et devinrent des proies plus faciles, alors qu'auparavant la tendance était inversée. La sélection directionnelle a ainsi changé la couleur de la population de phalènes du bouleau. De manière intéressante, un assainissement de l'air a plus tard ré-inversé la tendance et de nouveau favorisé les formes blanches.

ENCADRÉ 1.3

La **dérive génétique** est responsable de fluctuations dans les fréquences des variants d'une génération à l'autre par un effet d'échantillonnage aléatoire des gènes dans la population parentale pour créer la descendance (Charlesworth, 2009). Contrairement à la sélection, la dérive génétique agit indifféremment sur tous les variants. Cette force évolutive est très liée à la taille des populations, et plus précisément à la taille efficace de population (N_e). En 1931, Sewall Wright introduisait le concept de N_e , qu'il a défini comme la taille d'une population idéale, i.e. une population panmixtique^a de taille constante avec des générations qui ne se chevauchent pas, qui montrerait les mêmes variations dans les fréquences alléliques que dans la population étudiée (Wright, 1931). La taille efficace de population est souvent plus petite que la taille de population (Gutiérrez et al., 2012a; Vucetich et al., 1997). Elle peut être vue comme l'analogue évolutif de la taille totale ou recensée de population (Waples et al., 2014). Par exemple, la taille efficace de la population humaine a été estimée aux alentours de 10400 (Charlesworth, 2009). Plus N_e est petit, plus la dérive génétique sera forte. L'effet de la dérive génétique sur la composition d'une population est illustré dans la figure ci-dessous.

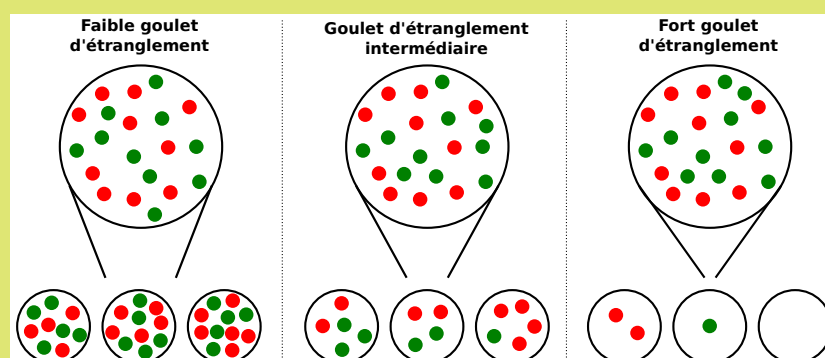


Illustration de l'effet de différentes tailles de goulets d'étranglement sur la composition de populations toutes issues de la même population parentale. Les trois grands cercles du haut représentent la même population parentale, composée d'un nombre équivalent de variants rouge et vert. Dans chaque cas de figure – goulet d'étranglement faible, intermédiaire ou fort – trois possibilités de populations issues de la population parentale sont représentées. Inspiré de Zwart and Elena (2015).

Des goulets d'étranglement allant de faible à fort sont associés à des N_e allant de grand à petit. Plus le goulet d'étranglement est étroit, plus N_e sera petit et plus la dérive génétique sera forte, affectant la composition génétique d'une population. En d'autres termes, plus la dérive génétique est forte, plus la composition génétique d'une population aura tendance à être différente d'une génération à la suivante. À l'extrême, de forts goulets d'étranglement peuvent conduire à la perte de variants (dans la figure cas de droite il ne reste que le variant rouge ou le variant vert), voire même à l'extinction de la population (dernier cas dans la figure).

a. Individus répartis de manière homogène et choix aléatoire des partenaires lors de la reproduction.

ENCADRÉ 1.4

La **migration** correspond à un flux de gènes ou de génotypes entre des populations d'agents pathogènes distinctes, typiquement entre plantes ou entre parcelles (McDonald and Linde, 2002a). Ce processus est responsable de la propagation des variants virulents dans le paysage agricole, voire même jusqu'à l'échelle planétaire par le vent (Brown and Hovmøller, 2002; Burdon, 1993).

Le scénario de contournement développé dans le paragraphe précédent n'en est en fait qu'un parmi d'autres. Il correspond typiquement à un cas où le variant *vir* apparaît dans une plante *S* dépourvue du gène majeur de résistance car dans notre exemple le variant *avr* est capable d'infecter la plante hôte et y est plus compétitif que le variant *vir* (Fig. 1.2, étapes 1 et 2). Au passage, on remarquera que puisque les deux variants *avr* et *vir* peuvent infecter la même plante, cela signifie que l'on se situe dans le cas du modèle gène-pour-gène (Table 1.1). Si le variant *vir* est initialement absent des cultures, il peut également arriver par immigration, depuis des plantes sauvages ou à cause de l'homme (Burdon, 1993; van den Bosch and Gilligan, 2003). Dans ce cas, s'il arrive directement sur une plante *R*, porteuse du gène majeur de résistance, il aura plus de chances de réussir l'infection et donc la deuxième étape, puisqu'il ne sera pas en compétition avec le variant *avr*. Cependant, il devra encore subir les effets de la stochasticité démographique (Encadré 1.5). Une autre possibilité est que le variant *vir* apparaisse dans une plante *R*. En effet, le variant *avr* arrive généralement à subsister et à se multiplier pendant quelques générations dans une plante *R* avant de s'éteindre; le modèle gène-pour-gène représente une situation extrême (Acosta-Leal and Xiong, 2008). Notamment, les résistances caractérisées par des réactions d'hypersensibilité, telle que celles procurées par le

gène de résistance *Tsw* chez le piment (*Capsicum chinense*) contre le *Tomato spotted wilt virus* (genre *Tospovirus*) (TSWV) (Moury et al., 1997), ou encore par le gène de résistance *L* chez le piment contre les tobamovirus (Kim et al., 2008), permettent une multiplication du variant *avr* avant que la plante ne se défende. En effet, dans ce cas la résistance de la plante n'est pas innée mais s'acquiert suite au contact avec le virus. Si par hasard le variant *avr* acquiert la ou les mutations de virulence avant que la plante n'ait réagi, le variant *vir* ainsi généré pourra infecter la plante s'il n'est pas perdu par dérive génétique.

ENCADRÉ 1.5

La **stochasticité démographique** correspond à des variations dans le taux de croissance d'une population ayant pour origine des différences aléatoires entre individus en terme de survie et de reproduction. Cela se traduit par des fluctuations de tailles de population d'une génération à l'autre par un effet d'échantillonnage dans la population des individus se reproduisant et de ceux ne se reproduisant pas. Ces fluctuations sont d'autant plus fortes que les populations sont de petite taille. À l'extrême, cela peut conduire à l'extinction de la population considérée, comme illustré dans la figure de l'encadré 1.3 dans le cas d'un fort goulet d'étranglement.

Au final, la dérive génétique relève de la génétique des populations alors que la stochasticité démographique relève de la dynamique des populations. Si en dynamique des populations on ne s'intéresse pas à la composition génétique des populations, la stochasticité démographique existe aussi dans les populations soumises à la dérive génétique.

Le variant *vir* est généralement affecté d'un *coût de fitness*, ou coût de virulence. C'est-à-dire que la ou les mutations lui ayant permis de devenir *vir*, ainsi que les mutations intermédiaires quand plusieurs sont nécessaires, peuvent entraîner une diminution de sa transmissibilité par vecteur ou par la graine, ainsi qu'une diminution de sa compétitivité dans la variété *S* voire dans des hôtes alternatifs (García-Arenal and McDonald, 2003; Moury et al., 2010). Ces coûts de fitness sont courants chez les virus de plantes (Sacristán and García-Arenal, 2008), où ils sont souvent élevés (Carrasco et al., 2007; Fraile et al., 2011). En effet, les génomes des virus sont de petite taille, souvent caractérisés par des régions codantes chevauchantes, et leurs protéines souvent multifonctionnelles. Ainsi, les mutations ne sont dans la plupart des cas pas neutres (Carrasco et al., 2007), et les mutations de virulence sont souvent associées des coûts de fitness (García-Arenal and McDonald, 2003).

1.2.2 Durabilité d'une résistance

La notion de contournement d'une résistance est intrinséquement liée à celle de durabilité : une résistance rapidement contournée n'est pas durable et inversement. Une résistance peut être très durable comme être contournée très facilement. Les résistances aux champignons et aux

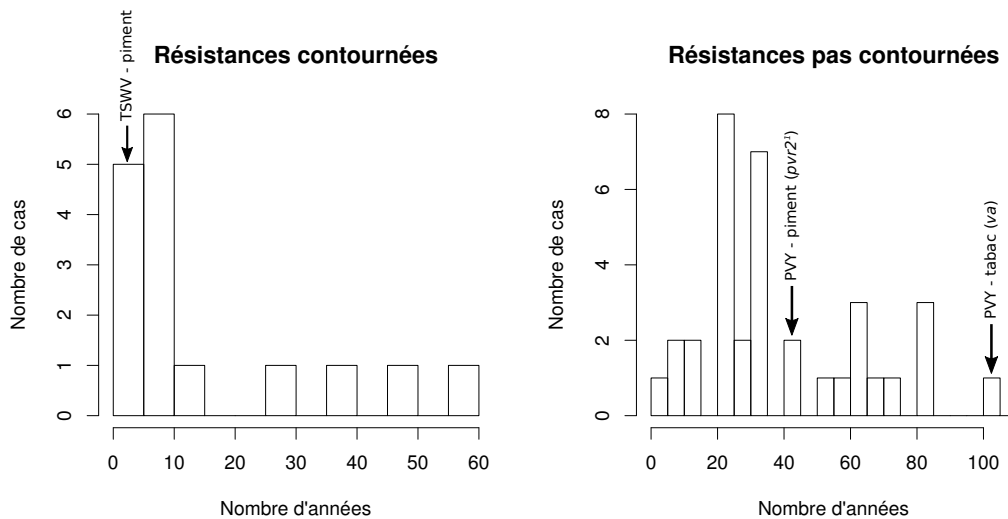


FIGURE 1.3 – Efficacité d’un ensemble de gènes de résistance aux virus issus de différents pathosystèmes et déployés dans les paysages agricoles. Les histogrammes représentent les distributions du nombre de cas, c’est-à-dire le nombre de couples pathosystèmes - localisation du déploiement, pour différentes classes de nombre d’années. L’histogramme de gauche correspond à des gènes de résistance qui sont déjà contournés, et l’histogramme de droite à des gènes de résistance qui n’ont pas encore été contournés (en 2003). Trois exemples, dont deux aux extrêmes, sont indiqués : (i) le pathosystème TSWV - piment, pour lequel une variété résistante déployée en Europe en 1999 a été contournée un an plus tard, (ii) le pathosystème PVY - piment (allèle majeur de résistance *pvr2¹*), pour lequel une variété résistante déployée en Europe à la fin des années 60 n’a pas encore été contournée, et (iii) le pathosystème PVY - tabac (allèle majeur de résistance *va*), pour lequel une variété résistante déployée en Europe au début du 20^{ème} siècle n’a toujours pas été contournée. Données issues de García-Arenal and McDonald (2003).

bactéries ne sont pas souvent durables alors que les résistances aux virus le sont davantage, même si tous les cas de figure ont été observés (Fig. 1.3) (García-Arenal and McDonald, 2003; Parlevliet, 2002). Un exemple flagrant de résistance non durable est la résistance *Rlm1* du colza (*Brassica napus*) au champignon *Leptosphaeria maculans* qui a été contournée en 3 ans seulement en France (Rouxel et al., 2003). En ce qui concerne les virus, de nombreux gènes de résistance déployés depuis plusieurs décennies montrent encore leur efficacité (Fig. 1.3, jusqu’à une centaine d’années) (García-Arenal and McDonald, 2003), tels que le gène *Tm2* chez la tomate et le gène *N* chez le tabac contre le *Tobacco mosaic virus* (genre *Tobamovirus*) (TMV), ou encore le gène *Me1* chez le piment contre le nématode *Meloidogyne incognita* (Castagnone-Sereno et al., 1996; Djian-Caporalino et al., 2011, 2014). À l’inverse, d’autres gènes de résistance ne se sont pas montrés durables, comme le gène *Tm1* chez la tomate contre le TMV (Parlevliet, 2002) ou le gène *Tsw* chez le piment contre le TSWV (Fig. 1.3).

La notion de durabilité d’une résistance est plutôt abstraite puisque difficilement quantifiable. Johnson a défini qu’une résistance était durable lorsqu’elle restait efficace suite à son déploiement sur une longue durée et à une grande échelle, dans un environnement favorable au

développement de l'agent pathogène (Johnson, 1984). Selon cette définition, la durabilité d'une résistance ne peut être identifiée qu'*a posteriori*, après son déploiement dans le paysage agricole sur le long terme (Fig. 1.2), et dans la plupart des cas, après son contournement (Leach et al., 2001). Des outils de prédiction de la durabilité *a priori* reposant sur la modélisation peuvent donc se révéler utiles dans ce contexte. Van den Bosch and Gilligan (2003) ont proposé et testé trois mesures de la durabilité d'une résistance grâce à un modèle épidémiologique simple. La première mesure, le temps au bout duquel l'agent pathogène *vir* envahit la population de plantes de la variété *R*, s'applique à un paysage dans lequel le variant *vir* est absent avant le déploiement de cette variété *R*. Cette mesure implique donc la réussite de deux processus : l'apparition du variant *vir*, par mutation ou immigration (pas de recombinaison dans leur modèle), puis son établissement dans la population d'agents pathogènes. Les deuxième et troisième mesures concernent un paysage dans lequel le variant *vir* est déjà présent en faible densité dans la population d'agents pathogènes avant le déploiement de la variété *R*. La deuxième mesure correspond au temps au bout duquel le variant *vir* 'prend le dessus', c'est-à-dire lorsqu'il atteint une fréquence prédéfinie (e.g. 0.9 dans leur étude). Enfin, la troisième mesure correspond au nombre d'hôtes sains cumulé dans le temps jusqu'à ce que le variant *vir* ait 'pris le dessus'. Dans leur étude, les auteurs mettent en avant des différences entre les trois mesures (Fig. 1.4), et pointent du doigt le fait que toutes les études précédentes ne se sont basées que sur une seule de ces mesures, à savoir la deuxième (temps au bout duquel le variant *vir* 'prend le dessus'). Notamment, la figure 1.4 montre les différences entre les formes de courbes de durabilité en fonction de la proportion de plantes *R* selon la mesure de durabilité choisie (en forme de U, de L, ou horizontales), mais aussi des effets contrastés des paramètres du modèle sur la variation des valeurs de durabilité (écart variable entre les courbes).

Les modèles mathématiques tels que celui de van den Bosch and Gilligan (2003) peuvent fournir des prédictions de durabilité à l'échelle du paysage agricole, mais elles seront plus qualitatives que quantitatives étant donné la nécessité de simplifier la réalité par la modélisation (Brauer and Castillo-Chávez, 2001). Du côté des biologistes, les mesures de durabilité sont généralement effectuées en laboratoire à une échelle plus petite, celle de l'interaction entre une plante et son agent pathogène. Cette échelle correspond aux deux premières étapes du contournement, à savoir l'apparition de variants *vir* et leur accumulation dans les plantes (Fig. 1.2). Dans un premier temps, différentes études ont permis une meilleure compréhension de la nature moléculaire de l'interaction entre les produits des gènes majeur de résistance chez les plantes et les facteurs d'avirulence chez les agents pathogènes, notamment grâce à

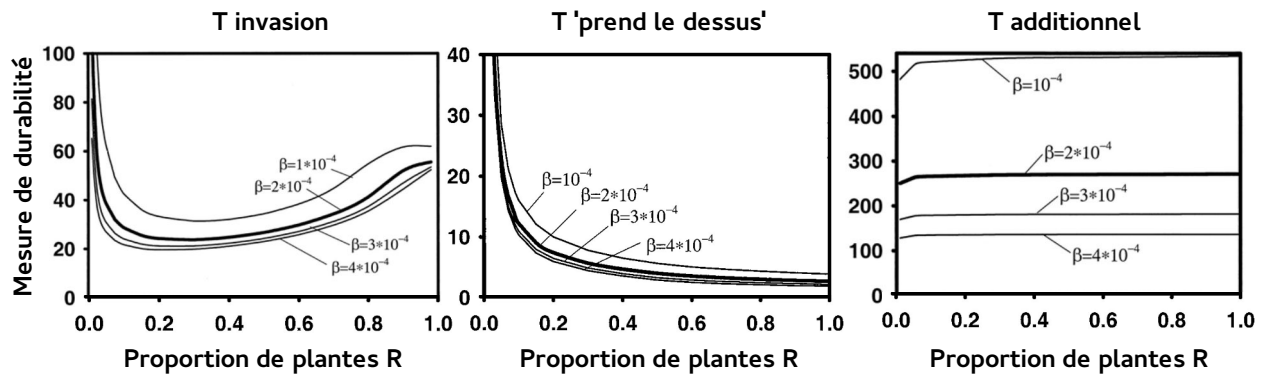


FIGURE 1.4 – Illustration des différences obtenues selon la mesure de durabilité choisie. Chaque graphique correspond à une mesure de durabilité : le temps T au bout duquel l’agent pathogène virulent *vir* envahit les plantes résistantes R (T invasion), le temps au bout duquel le variant *vir* ‘prend le dessus’ (T ‘prend le dessus’) et le nombre d’hôtes sains cumulé dans le temps jusqu’à ce que le variant *vir* ait ‘pris le dessus’ (T additionnel). Pour chaque graphique, les valeurs de durabilités sont représentées en ordonnée en fonction de la proportion de plantes R , et pour différentes valeurs du taux d’infection β . Créée à partir de van den Bosch and Gilligan (2003).

des échantillonnages de la diversité existante chez les pathogènes et leurs plantes (e.g. Gómez et al., 2009; Hammond-Kosack and Parker, 2003; Harrison, 2002; Kang et al., 2005; Robaglia and Caranta, 2006; Rouxel and Balesdent, 2010). Dans un second temps, différentes études se sont attachées à mesurer la durabilité de gènes majeurs de résistance aux agents pathogènes, ou d’établir un critère de risque de contournement de ces gènes de résistance, grâce à différentes approches.

Une première approche consiste à effectuer des échantillonnages au champ, chez les agriculteurs. Par exemple, McDonald and Linde (2002b) rapportent différents champignons infectant des variétés de céréales, identifiés la plupart du temps dans des champs d’agriculteurs. Ils attribuent à chaque espèce de champignon un potentiel évolutif pour le contournement d’un gène majeur de résistance, basé sur l’analyse de l’intensité de cinq forces évolutives : mutation, dérive génétique, sélection, migration et recombinaison. Enfin, ils proposent un diagramme de décision pour aider à la création et au déploiement de variétés résistantes qui devraient être plus durables. Les décisions dépendent des caractéristiques des agents pathogènes, notamment de leur mode de reproduction, de l’ordre de grandeur de leurs tailles efficaces de populations et de l’intensité des flux migratoires.

Une seconde approche se base sur l’évolution expérimentale au champ. Cette approche est basée sur l’inoculation d’un variant *avr* à des plantes porteuses d’un gène majeur de résistance. L’expérience étant effectuée au champ, la première étape du contournement (mutation vers le variant *vir*) n’est pas forcément bien appréhendée car un variant *vir* peut aussi venir

de l'extérieur vers les cultures. Par exemple, Brun et al. (2010) ont réalisé une expérience d'évolution au champ sur 5 ans, en inoculant une population locale de champignons *Leptosphaeria maculans* à des plantes de colza (*Brassica napus*) porteuses du gène majeur de résistance *Rlm6*, puis ont évalué la durabilité de la résistance grâce à des mesures de pourcentages de plantes présentant des symptômes révélateurs d'une infection (lésions et chancres de la tige).

Enfin, une troisième approche se base sur l'évolution expérimentale en conditions contrôlées. Avec cette approche, les troisième et quatrième étapes du contournement (dissémination à l'échelle du paysage et saisonnalité) ne sont généralement pas prises en compte. Par exemple, Palloix et al. (2009), puis Quenouille et al. (2013, 2014) se sont intéressés à la durabilité du gène majeur de résistance *pvr2³* du piment (*Capsicum annuum*) au PVY, qu'ils ont évalué grâce au pourcentage de plantes infectées au niveau systémique un mois après l'inoculation avec un variant *avr*. Barbary et al. (2014) ont étudié la durabilité de deux gènes majeurs de résistance du piment (*Capsicum annuum* L.) aux nématodes *Meloidogyne incognita* en inoculant un variant *avr* du nématode à très forte concentration puis en calculant le pourcentage de plantes présentant plus de 5 kystes, seuil à partir duquel ils considèrent la résistance contournée. À l'intersection entre évolution expérimentale au champ et en conditions contrôlées, Fournet et al. (2013) ont fait évoluer des nématodes (*Globodera pallida*) provenant de champs infestés par passages successifs sur des variétés de pommes de terre (*Solanum* spp.) porteuses de la même résistance *GpaV*. Pour évaluer la durabilité du gène *GpaV*, ils ont comparé la fréquence du variant *vir* avant et après leur expérience de passages successifs.

Nous avons pu voir différentes approches théoriques et biologiques pour mesurer la durabilité d'un gène majeur de résistance à un agent pathogène, ainsi que la gamme de variation des durabilités mesurées dans les agroécosystèmes après déploiement d'une variété résistante (Fig. 1.3). Nous allons maintenant discuter de l'impact de la pression de sélection exercée sur l'agent pathogène lors du déploiement d'une variété résistante sur l'évolution de cet agent pathogène, et du risque de la monoculture sur l'émergence rapide de variants pathogènes virulents.

1.2.3 Monocultures et risques d'épidémies

Les résistances qualitatives qui suivent le modèle gène-pour-gène induisent une forte pression de sélection sur les populations d'agents pathogènes (Moury et al., 2010; Zhan et al., 2015). En effet, le variant *avr* ne pouvant pas infecter les plantes *R* ou seulement de manière très ponctuelle et limitée, le premier variant *vir* apparu et transmis à une plante *R* a de

forte chances de conduire rapidement à une épidémie à grande échelle. Brown et al. (2002) explique ce phénomène grâce à la métaphore de l'Oasis dans le Désert, correspondant à un contexte où la majorité des parcelles sont cultivées avec une variété R . Ainsi, du point de vue de l'agent pathogène, initialement *avr*, les parcelles de plantes S constituent de petites surfaces d'environnement favorable parmi de grandes surfaces inhabitables (parcelles de plantes R). Le premier mutant capable d'infecter les plantes R (variant *vir*) a un très fort avantage sélectif, et une large étendue d'habitat favorable disponible s'offre à lui. Suite à sa dissémination, la résistance sera contournée et deviendra inefficace. À l'échelle du paysage, les monocultures sans rotation⁴ induisent donc une forte pression de sélection sur les agents pathogènes et peuvent conduire à des épidémies majeures (Fig. 1.5, Zhan et al. (2015)). En effet, si l'agent pathogène arrive à s'adapter et à infecter une première plante R cultivée, la suite sera d'autant plus facile pour lui qu'il aura à sa disposition des parcelles entières de plantes R . Le risque épidémique n'est pas toujours plus fort dans les monocultures, mais dépend des caractéristiques du paysage et du pathosystème étudié. Notamment, van den Bosch and Gilligan (2003) et Fabre et al. (2012b) ont montré que dans certains cas la durabilité d'une résistance était plus forte lorsque seule la variété R était cultivée (voir Fig. 1.4, graphique de gauche : si l'on veut déployer plus de 20% de plantes résistantes, il vaut mieux en déployer 100%). À l'inverse, procéder à des rotations dans les cultures peut permettre de protéger la résistance face à l'agent pathogène en augmentant le degré d'hétérogénéité de la sélection des variants *vir* et *avr* (Fabre et al., 2015; Zhan et al., 2015).

La constante évolution de l'agent pathogène pour s'adapter à la résistance de l'hôte a souvent été mise en relation avec la notion de cycles expansion-récession (*boom-and-bust cycles*) (Brown, 2015; Johnson, 1961; McDonald and Linde, 2002a; Zhan et al., 2015). Dans le contexte d'un paysage agricole, la phase d'expansion correspond au déploiement d'un gène majeur de résistance de grande efficacité à une échelle importante. Là où les populations d'agents pathogènes ciblées finiront très probablement par s'adapter à ce gène majeur grâce à l'apparition de variants *vir* qui vont être disséminés dans le paysage, anéantissant l'efficacité de la résistance (phase de récession) (McDonald and Linde, 2002a,b). Des scénarios en cycles d'expansion-récession ont été reportés à plusieurs reprises dans les cultures, notamment chez les rouilles ainsi que chez les oïdiums des céréales (McDonald and Linde, 2002a). Un exemple d'actualité concerne le déploiement du gène *Rpi-blb1* contre le mildiou de la pomme de terre (*Phytophthora infestans*), qui était efficace contre toutes les souches connues de cet agent pathogène (Brown, 2015; Song et al., 2003; van der Vossen et al., 2003). Des souches virulentes vis-à-vis de *Rpi-blb1*

4. Succession de cultures de différents types dans le temps dans une même parcelle.

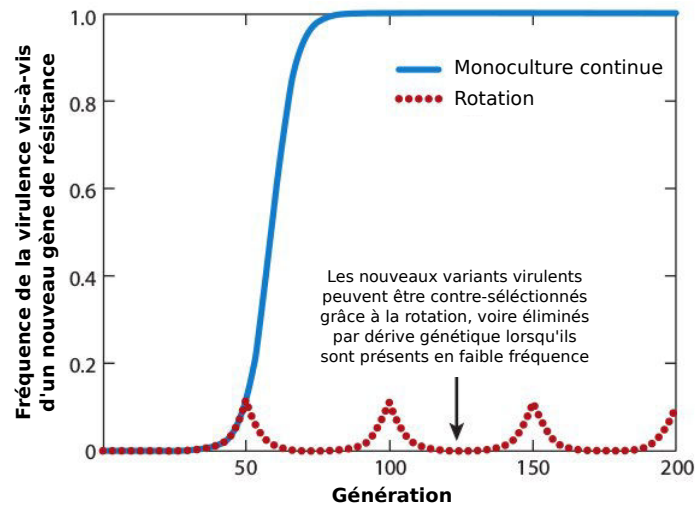


FIGURE 1.5 – Évolution attendue d'un agent pathogène confronté à des plantes porteuses d'un nouveau gène de résistance sous des conditions de monoculture (courbe bleue), ou avec une rotation des cultures (courbe en pointillés rouges), sans possibilité de recombinaison. Modifié d'après Zhan et al. (2015).

ont été découvertes au Mexique (Champouret et al., 2009) et aux Pays-Bas (Li et al., 2012). Bien que *Rpi-blb1* soit encore efficace contre la plupart des souches de *P. infestans* et qu'il soit utilisé en sélection pour la pomme de terre, la phase de récession a commencé et la durabilité du gène *Rpi-blb1* va dépendre du coût de fitness associé aux mutations de virulence et des stratégies de gestion de la résistance employées (Brown, 2015). Une autre étude a été rapportée sur le pathosystème tobamovirus - piment, impliquant des relevés au champ entre 1984 et 2004, lors des épidémies de tobamovirus (Fraile et al., 2011). Pour cinq type de variétés de piment, différant par l'allèle présent au locus de résistance *L*, déployés en différentes proportions, la réponse des populations de tobamovirus a été analysée. Les auteurs ont montré un effet des allèles de résistance sur l'évolution des tobamovirus, ainsi que des coûts de fitness associés à la virulence.

Le cycle expansion-récession est illustré figure 1.5. À la génération 0, la variété *R* est déployée et la fréquence des variants *vir* est proche de 0. Un peu avant la cinquantième génération, les variants *vir* commencent à envahir la population d'agents pathogènes, et arrivent très rapidement, avant la centième génération, à une fréquence de 1. À l'inverse, la rotation des cultures permet de relâcher la pression de sélection sur les variants *vir* ; ainsi ils n'envahissent jamais entièrement la population d'agents pathogènes et la résistance n'est pas contournée (courbe en pointillés rouges, Fig. 1.5).

Différents modèles théoriques et expériences ont pu démontrer l'effet bénéfique de différentes stratégies de gestion des résistances sur la diminution du risque épidémique. Alors que le

déploiement d'un seul gène de résistance en monoculture dirige sur le long terme la sélection des populations d'agents pathogènes vers la virulence, différentes stratégies peuvent modifier la manière dont la sélection opère sur les populations d'agents pathogènes (McDonald and Linde, 2002a). Parmi celles-ci figure le mélange de variétés, qui aurait un effet bénéfique via l'effet de dilution. Ce concept correspond au maintien de variants *avr* dans la population d'agents pathogènes grâce à la présence de plantes *S* dans lesquelles les variants *vir* sont généralement défavorisés, à cause du coût de fitness des mutations de virulence (Fabre et al., 2012b; Gilligan, 2008; Mundt, 2002; Ostfeld and Keesing, 2012; Wolfe, 1985). Plus généralement, la rotation des cultures dans le temps et leur mélange dans l'espace peut apporter un effet bénéfique en générant de la sélection disruptive (Fig. 1.5) (Fabre et al., 2015; REX Consortium, 2013; Zhan et al., 2015), qui favorise différents mutants en différents lieux et/ou à différents moments (McDonald and Linde, 2002a). Le pyramidage de plusieurs gènes de résistance constitue aussi une stratégie efficace, l'agent pathogène étant souvent difficilement capable d'enchaîner la séquence de mutations correspondant au contournement de chaque gène de résistance (McDonald and Linde, 2002a).

1.2.4 Un enjeu majeur : intégrer les principes éco-évolutifs régissant les interactions hôtes - agents pathogènes pour une meilleure gestion des cultures.

Une stratégie qui est évoquée de plus en plus souvent dans la littérature consiste à jouer sur les forces évolutives pour gérer le déploiement des résistances et favoriser leur durabilité (Zhan et al., 2015). En effet, la compréhension des processus conduisant au contournement d'une résistance passe par la compréhension des processus qui gouvernent l'évolution des agents pathogènes (McDonald and Linde, 2002a). Cette évolution est complexe puisqu'elle est liée à l'ensemble de l'écosystème associé à l'agent pathogène, incluant notamment les plantes hôtes, et pour certains agents pathogènes tels que les virus leurs insectes vecteurs (Dedryver et al., 2010). L'évolution est le produit des différentes forces évolutives agissant sur les populations d'agents pathogènes et de leurs interactions (McDonald and Linde, 2002a). Étudier la variabilité de la composition génétique des populations d'agents pathogènes dans leurs plantes hôtes et ses changements sont les clés de la compréhension des forces évolutives qui les gouvernent et peut permettre de développer des stratégies adaptées pour le contrôle des maladies des plantes (García-Arenal et al., 2001).

Les forces guidant l'évolution sont universelles et agissent donc de la même manière dans

les écosystèmes naturels que dans les milieux cultivés (Burdon, 1993). Les interactions entre les plantes sauvages et leurs agents pathogènes se produisent depuis des millénaires, et les processus co-évolutifs ont la plupart du temps naturellement conduit à une limitation des épidémies dans le temps et dans l'espace (Burdon and Thrall, 2014; Zhan et al., 2015). Cette limitation spatio-temporelle est due à la longue histoire de co-évolution entre les hôtes et leurs agents pathogènes, conduisant les populations d'hôtes à être plutôt distribués en patches, et à présenter une grande diversité intra- et inter-populations en termes de résistances génétiques (Laine et al., 2011; Zhan et al., 2015). L'agriculture moderne a perturbé ces interactions notamment via la monoculture et l'utilisation à grande échelle de variétés porteuses de résistances très efficaces. Ainsi, nous avons créé des situations où la pression de sélection est très forte sur l'agent pathogène et risque de conduire à son évolution rapide et à des épidémies majeures (Fig. 1.6) (Stukenbrock and McDonald, 2008).

Stukenbrock and McDonald (2008) et Elena et al. (2014) mettent en avant quatre facteurs ayant un rôle majeur sur l'émergence des maladies dans les agroécosystèmes : (i) la diversité réduite des espèces de plantes par rapport aux systèmes naturels, (ii) la densité d'hôtes plus élevée, (iii) la diversité génétique réduite au sein d'une culture comparée à celle des populations sauvages d'une même espèce ou apparentées, et (iv) une plus forte homogénéité environnementale (Fig. 1.6). Notamment, plusieurs études se sont concentrées sur la relation entre la biodiversité et le risque de maladie (Keesing et al., 2010). Une étude récente a été réalisée sur le piment sauvage (*Capsicum annuum* var. *glabriusculum*) au Mexique (Pagán et al., 2012). Trois types de populations de piment correspondant à des degrés d'intervention humaine croissants ont été analysées. Ainsi, une intervention humaine accrue a été associée à une réduction de la diversité génétique des populations de piment sauvage (González-Jara et al., 2011) et à une augmentation du risque d'infection virale et de maladie symptomatique (Pagán et al., 2012). Plus généralement, Stukenbrock and McDonald (2008) affirment que les caractéristiques propres aux agroécosystèmes facilitent la transmission des agents pathogènes entre plantes infectées et saines, et créent un environnement plus propice à la spécialisation à un hôte et à l'émergence de variants virulents que les systèmes naturels.

De plus, l'agriculture occulte la possibilité d'une évolution des plantes puisque les variétés cultivées sont sélectionnées par l'homme. Ce n'est donc plus une co-évolution que l'on étudie dans les systèmes agricoles, mais une évolution des agents pathogènes uniquement. Plus précisément, ce sont les contraintes et opportunités écologiques, qui sont très différentes entre les systèmes agricoles et naturels, qui ont des effets substantiels sur les dynamiques de populations d'agents pathogènes et ont ainsi des conséquences épidémiologiques et génétiques très différentes



FIGURE 1.6 – Illustration des différences principales entre les écosystèmes naturel et agricole. Les agroécosystèmes (figure du bas) montrent une plus forte homogénéité environnementale, une plus faible diversité des espèces, et une plus forte densité d’hôtes qui facilitent la transmission des agents pathogènes à des plantes voisines saines. Ces caractéristiques rendent possible le développement d’épidémies conduisant à l’émergence d’agents pathogènes virulents et/ou très agressifs, et spécialisés sur une ou un faible nombre d’espèce(s) d’hôtes. Les écosystèmes naturels (figure du haut) montrent une plus forte hétérogénéité environnementale, une plus forte diversité des espèces, et une plus faible densité d’hôtes qui favorisent des agents pathogènes avirulents et/ou moins agressifs et généralistes, pouvant infecter diverses espèces d’hôtes. Source : Stukenbrock and McDonald (2008).

dans ces deux environnements. Comme évoqué précédemment, les différences écologiques les plus notables concernant les populations de plantes hôtes entre les deux milieux sont les tailles de population, leur densité et leur distribution spatiale, leur variabilité génétique et leur continuité dans le temps. Chez certains agents pathogènes, une taille minimale de population d'hôtes est nécessaire au maintien de ces derniers, condition plus facilement remplie dans les cultures. La densité et la distribution spatiale des hôtes est à mettre en relation avec le mode de dispersion des agents pathogènes (e.g. par le sol, par des vecteurs, par l'air). Selon ces capacité et distance de dispersion, l'agent pathogène sera favorisé dans certaines configurations spatiales de sa population d'hôtes mais pas dans d'autres. Dans les cultures, les grandes surfaces généralement impliquées ainsi que les fortes densités d'hôtes ne pénalisent aucun mode de dispersion. La relation entre la densité des hôtes et l'incidence d'une maladie peut varier au sein d'un même pathosystème et d'un pathosystème à un autre. Cette relation est généralement positive, mais parfois seulement à court terme. La variabilité génétique des populations d'hôtes est sans aucun doute beaucoup plus importante dans les écosystèmes naturels. Dans les systèmes agricoles, le mélange de variétés, comparé à la monoculture conduit à une diminution des tailles de population d'agents pathogènes, mais aussi à une augmentation de la diversité génétique de ces derniers. Enfin, des variations annuelles de la présence d'hôtes auront très probablement des conséquences sur le taux de croissance et de reproduction de l'agent pathogène (Burdon, 1993).

Concernant les systèmes agricoles, McDonald and Linde (2002a) ont émis l'hypothèse que le potentiel évolutif des espèces d'agents pathogènes – champignons ou bactéries – était en lien avec la structure génétique de leurs populations. Selon cette théorie, les agents pathogènes présentant le plus fort risque de contournement des résistances sont ceux à reproduction mixte (i.e. sexuée et asexuée), avec un fort taux de migration entre populations, des grandes tailles efficaces de population et des taux de mutation importants. À l'inverse, les agents pathogènes à reproduction asexuée, avec un faible taux de migration entre populations, de petites tailles efficaces de population et des taux de mutation faibles présentent un faible risque de contournement. García-Arenal and McDonald (2003) ont repris ce concept de potentiel évolutif pour les virus de plantes. Ces auteurs ont estimé le potentiel évolutif en se basant sur trois facteurs évolutifs : (i) la taille efficace des populations, (ii) les flux de gènes et de génotypes, et (iii) le système reproductif. Leur étude montre que ce potentiel évolutif peut constituer un facteur déterminant pour la durabilité d'un gène majeur de résistance aux virus de plantes.

Au final, intégrer les principes éco-évolutifs régissant les interactions entre les hôtes et leurs agents pathogènes dans les milieux cultivés pourrait permettre de minimiser le potentiel évolutif des agents pathogènes. Cela conduirait à une gestion plus durable des résistances, et à

une meilleure productivité. Différentes approches peuvent permettre d’y arriver. Par exemple, la dérive génétique, en réduisant la variabilité génétique des populations d’agents pathogènes, peut permettre de réduire leur capacité d’évolution étant donné que les variants *vir* doivent normalement être présents en faible fréquence avant le déploiement d’une variété résistante. Autre exemple, la réduction des flux de gènes entre les populations d’agents pathogènes présentes dans chaque plante infectée peut limiter la propagation de variants *vir*. L’éradication d’hôtes alternatifs peut minimiser les sources d’inoculum primaire (inoculum conduisant à la première infection dans le milieu cultivé). Enfin, favoriser la sélection de divers mutants pathogènes (sélection diversifiante) grâce au mélange de variétés où à leur alternance dans le temps peut permettre de réduire la pression de sélection sur les agents pathogènes (Zhan et al., 2015). Parmi ces approches, celle favorisant la dérive génétique m’intéresse tout particulièrement. Je vais maintenant présenter un deuxième type de résistance des plantes et voir en quoi cette résistance pourra nous permettre d’exploiter la dérive génétique afin de retarder l’émergence de variants pathogènes adaptés à la résistance qualitative des plantes.

1.3 La résistance quantitative comme moyen de protection d’une résistance qualitative

L’interaction entre une plante et un agent pathogène peut être en partie déterminée par des traits quantitatifs qui peuvent prendre un continuum de valeurs selon les variants de l’agent pathogène, les génotypes de plante et l’environnement (Pariaud et al., 2009). Ces traits quantitatifs sont contrôlés génétiquement chez l’hôte et l’agent pathogène par des loci à effet quantitatif (*Quantitative Trait Loci*) (QTL) présents dans leur fond génétique (Fig. 1.7) (Lannou, 2012). Du côté des plantes, la cartographie de ces QTL dans leur fond génétique permet de caractériser les résistances *quantitatives* ou partielles aux agents pathogènes. Les traits quantitatifs mesurés sont souvent liés à l’agressivité de l’agent pathogène et ont été principalement mesurés dans le cas des interactions avec les champignons et les oomycètes. La majorité de ces traits concernaient l’efficacité d’infection (probabilité qu’une spore déposée sur un hôte produise une lésion), la période de latence (intervalle de temps entre l’infection et les premières productions de spores provenant de cette infection), la période infectieuse (intervalle de temps entre le début et la fin de la sporulation), la taille des lésions (surface de production des spores) (Funnell et al., 2001), le taux de production des spores (quantité de spores produites

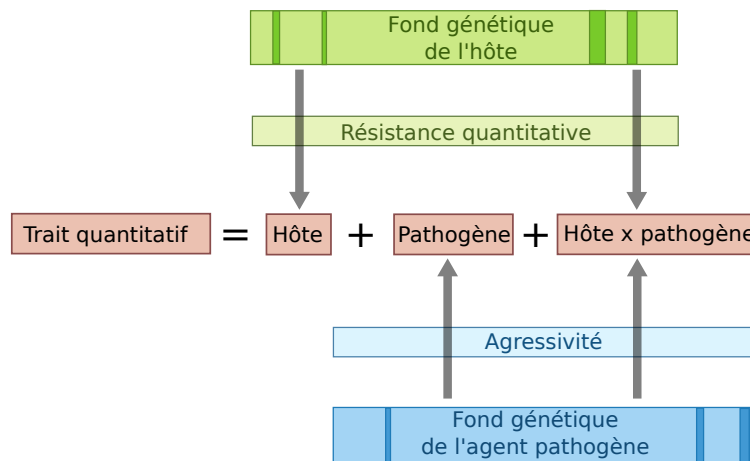


FIGURE 1.7 – Un trait quantitatif d’agressivité peut être représenté comme résultant d’un effet de l’hôte, de l’agent pathogène ainsi que de leur interaction (en rouge). Le fond génétique de l’hôte, dans lequel des QTL pour une résistance (en vert) peuvent être détectés, détermine le niveau de la résistance quantitative (effet de l’hôte et interaction). Le fond génétique de l’agent pathogène, dans lequel des QTL d’agressivité (en bleu) peuvent être détectés, détermine le niveau d’agressivité de l’agent pathogène (effet de l’agent pathogène et interaction). Selon le trait considéré et les génotypes de l’hôte et de l’agent pathogène testés, certains effets peuvent être absents. Modifié d’après (Lannou, 2012).

par lésion et par unité de temps) et la production de mycotoxines⁵ (qui induisent des nécroses⁶, composantes de l’agressivité chez certains agents pathogènes) (Lannou, 2012; Pariaud et al., 2009).

La première publication sur l’identification de QTL contrôlant une résistance chez les plantes concernait le pathosystème riz - *Magnaporthe grisea* (agent responsable de la pyriculariose) (Wang et al., 1994). Depuis, de nombreuses autres détections ont été publiées, notamment pour des maladies à champignons, telles que d’autres détections pour la pyriculariose du riz (Ballini et al., 2008), des détections pour la tavelure du pommier (*Venturia inaequalis*) (Calenge et al., 2004) ou encore pour la rouille brune de l’orge (*Puccinia hordei*) (Marcel et al., 2008).

Du côté des agents pathogènes, peu d’études se sont attachées à la cartographie de QTL impliqués dans le pouvoir pathogène dans leur fond génétique (Pariaud et al., 2009). Récemment, Doumayrou et al. (2015) s’y sont intéressés, identifiant un QTL chez les virus (recombinants entre deux espèces de begomovirus) impliqués dans l’accumulation virale intra-hôte et la précocité d’infection. Une étude plus ancienne a montré que le fond génétique du PVY pouvait modifier la fréquence de contournement du gène majeur de résistance *pvr2³* chez le piment (Montarry et al., 2011). En ce qui concerne les champignons et les bactéries, différentes études ont montré une variation significative d’un ou de plusieurs traits d’agressivité au sein

5. Composé toxique produit par un champignon (D’Arcy et al., 2001).

6. Mort de cellules ou tissus chez la plante (D’Arcy et al., 2001).

d'une population d'agents pathogènes appartenant au même pathotype⁷ ou partageant des génotypes similaires (Pariaud et al., 2009). Les traits étudiés concernaient l'efficacité d'infection, la taille des lésions, la capacité de sporulation ou encore la période de latence. Des variations allant jusqu'à 91% entre les isolats les moins et les plus agressifs ont été reportées chez *Cochliobolus carbonum*. Pariaud et al. (2009) rapportent deux études ayant identifié des QTL de taille de lésion ou de croissance du champignon chez *Gibberella zeae* (Cumagun et al., 2004) et chez *Heterobasidion annosum* (Lind et al., 2007).

Au final, l'expression d'un trait quantitatif peut dépendre de l'hôte, de l'agent pathogène ainsi que de leur interaction (Fig. 1.7) (Lannou, 2012). Comme nous l'avons vu, dans un contexte agronomique, la plupart des QTL de plantes identifiés permettent le contrôle d'un ou de plusieurs traits quantitatifs de l'agent pathogène.

1.3.1 Le fond génétique de la plante peut moduler le taux de contournement d'un gène majeur de résistance

Le pyramidage de gènes majeurs de résistance chez les plantes présente des avantages et des inconvénients (Pink, 2002). Notamment, le risque majeur du déploiement de gènes majeurs pyramidés vient de la forte pression sélection unidirectionnelle agissant sur l'agent pathogène, le poussant à s'adapter à l'ensemble des résistances qualitatives pyramidées pour devenir virulent, et donc la perte de toutes les résistances qualitatives simultanément. Récemment, plusieurs études ont montré l'intérêt d'une stratégie de pyramidage alternative et prometteuse. Celle-ci consiste à combiner des résistances quantitatives avec une résistance qualitative chez les plantes pour augmenter la durabilité de la résistance qualitative (Brun et al., 2010; Fournet et al., 2013; Palloix et al., 2009). Brun et al. (2010) se sont concentrés sur le pathosystème colza (*Brassica napus*) - champignon (*Leptosphaeria maculans*), au champ. Ils ont observé que le gène de résistance qualitative *Rlm6* était contournée en trois saisons de culture lorsqu'il est introgressé dans un fond génétique sensible, contre cinq ans dans un fond génétique résistant.

Fournet et al. (2013) ont montré sur le pathosystème pomme de terre (*Solanum* spp.) - nématode (*Globodera pallida*) que le fond génétique de l'hôte avait un impact sur l'adaptation au QTL majeur de résistance *GpaV*, et ont suggéré que la présence de résistances quantitatives dans ce fond génétique pouvait expliquer une plus grande durabilité de *GpaV*.

Enfin, Palloix et al. (2009) se sont intéressés aux fréquences de contournement de l'allèle

7. Subdivision d'une espèce d'agent pathogène selon leur virulence ou avirulence pour un ensemble de variétés d'hôtes différentes.

majeur de résistance *pvr2³* au PVY chez le piment pour deux variétés de plantes, une variété avec l'allèle majeur de résistance *pvr2³* introgressé dans un fond génétique sensible 'Rs' et une variété avec l'allèle majeur de résistance *pvr2³* introgressé dans un fond génétique avec des résistances quantitatives 'Rr'. Suite à l'inoculation des deux variétés Rs et Rr avec un variant avirulent de PVY (clone SON41p), ils ont observé en moyenne 27.6% d'infection systémique dans les plantes Rs et 0% dans les plantes Rr.

Tous ces résultats illustrent bien l'intérêt des résistances quantitatives pour améliorer le contrôle d'une maladie et augmenter la durabilité d'une résistance qualitative. Dans la suite du document, j'appellerai les combinaisons de résistances qualitative et quantitative au sein d'un même génotype d'hôte des résistances *polygéniques*.

Je vais me focaliser sur le dernier pathosystème évoqué, à savoir l'interaction piment (*Capsicum annuum*) - PVY, modèle d'étude de cette thèse. Quenouille et al. (2014) ont tenté d'apporter des éclaircissements aux résultats de Palloix et al. (2009), en cartografiant des QTL liés à la durabilité de l'allèle majeur de résistance *pvr2³* ou liés à des résistances quantitatives. Une collection de 350 lignées haploïdes doublées (HD) de piments a été obtenue à partir du croisement entre une variété sensible (Yolo Wonder), porteuse de l'allèle de sensibilité *pvr2⁺*, et une variété résistante (Perennial), porteuse de l'allèle de résistance *pvr2³* au même locus (Fig. 1.8) (Caranta et al., 1997). Deux sous-populations ont été définies à partir de cette collection, l'une porteuse de l'allèle de sensibilité *pvr2⁺* et l'autre porteuse de l'allèle de résistance *pvr2³*. Au sein de chaque sous-population, les différences résident donc dans le fond génétique. Les QTL identifiés par Quenouille et al. (2014) dans la sous-population portant l'allèle de résistance *pvr2³* concernent l'accumulation de l'agent pathogène, l'aire sous la courbe de progression des symptômes et la fréquence de contournement de *pvr2³* (Fig. 1.9). Le but de l'étude de Quenouille et al. (2014) était d'identifier des facteurs génétiques chez la plante modulant la fréquence d'adaptation au gène majeur de résistance *pvr2³*, et de les comparer à ceux contrôlant la résistance quantitative, en l'occurrence liées à la sévérité des symptômes et à l'accumulation virale. La fréquence de contournement du gène majeur *pvr2³* a permis la détection de trois QTL. Du côté des résistances quantitatives, trois QTL ont été identifiés pour l'aire sous la courbe de progression des symptômes (*area under the disease progress curve*, AUDPC), et deux QTL pour la concentration virale relative entre lignées HD. Les résultats de cette étude montrent que trois des quatre QTL modulant la fréquence de contournement de *pvr2³* colocalisent avec les QTL modulant les résistances quantitatives (accumulation virale et/ou sévérité des symptômes). Si l'on considère uniquement ces QTL qui colocalisent, les allèles

qui augmentent la résistance quantitative diminuent la fréquence de contournement. Ainsi, les auteurs suggèrent que ces colocalisations pourraient être dues à un effet pléiotropique⁸ des facteurs de résistances quantitatives sur la fréquence de contournement de *pvr2*³.

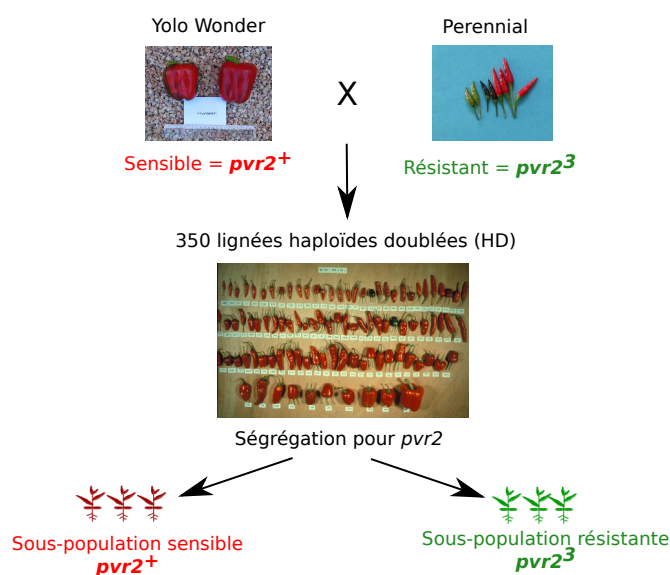


FIGURE 1.8 – Lignées haploïdes doublées (HD) étudiées dans le cadre de cette thèse. Les 350 lignées HD sont issues de l’hybride F₁ entre les génotypes de *Capsicum annuum* Yolo Wonder (parent sensible, porteur de l’allèle de sensibilité *pvr2*⁺) et Perennial (parent résistant, porteur de l’allèle de résistance *pvr2*³). Une ségrégation pour le locus du gène majeur *pvr2* assistée par marqueurs génétiques a permis de différencier une sous-population porteuse de l’allèle de sensibilité et une autre sous-population porteuse de l’allèle de résistance. Dans le cadre de cette thèse je me suis intéressée à la sous-population porteuse de l’allèle de résistance *pvr2*³.

8. Gène ou mutation déterminant plusieurs caractères phénotypiques.

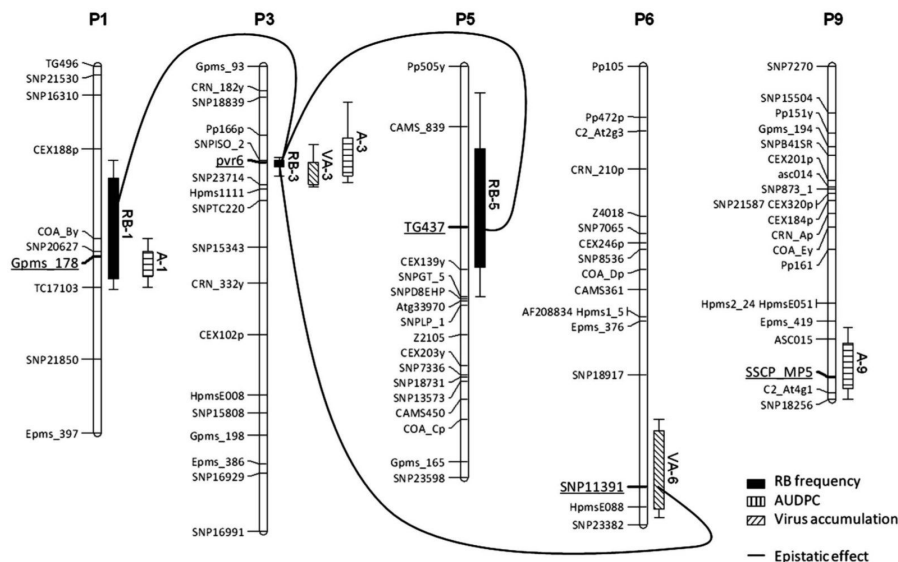


FIGURE 1.9 – Cartographie génétique des loci à effet quantitatif (*Quantitative Trait Loci*) (QTL) concernant le pathosystème piment (*Capsicum annuum*) - PVY. La population de lignées haploïdes doublées (HD) de piment utilisée est issue du croisement entre une variété sensible (Yolo Wonder) et une variété résistante (Perennial). Les traits quantitatifs analysés sont la fréquence de contournement du gène majeur *pvr2³* (RB frequency), l'aire sous la courbe de progression des symptômes (AUDPC) et l'accumulation virale (virus accumulation). Pour chaque QTL, la longueur des boîtes représente l'intervalle de confiance (IC) du QTL obtenu avec le logiciel MCQTL. Les lignes indiquent les effets épistatiques entre loci. Source : (Quenouille et al., 2014).

1.3.2 Trois mécanismes pour expliquer la plus grande durabilité de la résistance polygénique

Quenouille et al. (2013, 2014) ont essayé de comprendre quels mécanismes évolutifs pouvaient expliquer l'effet protecteur du fond génétique des plantes de piment vis-à-vis de la durabilité du gène majeur *pvr2³* Palloix et al. (2009). Je vais ici présenter les hypothèses et résultats de l'article de Quenouille et al. (2013) qui est à l'origine des questions que j'ai explorées pendant ma thèse.

Réduction de l'accumulation du virus

La première hypothèse concerne un lien entre la fréquence de contournement de l'allèle majeur de résistance *pvr2³* et l'accumulation du PVY. Ces deux traits ont été mesurés chez quatorze lignées HD porteuses de *pvr2³*, inoculées avec un variant avirulent de PVY pour la mesure de la fréquence de contournement, et avec un variant virulent pour la mesure d'accumulation virale. Un lien très significatif entre ces deux variables ($p_{value} < 10^{-5}$) et un bon ajustement entre les fréquences de contournement prédites et observées (R^2 de McFadden = 0.79) (Fig. 1.10) ont été rapportés. La colocalisation de certains des QTL contrôlant ces

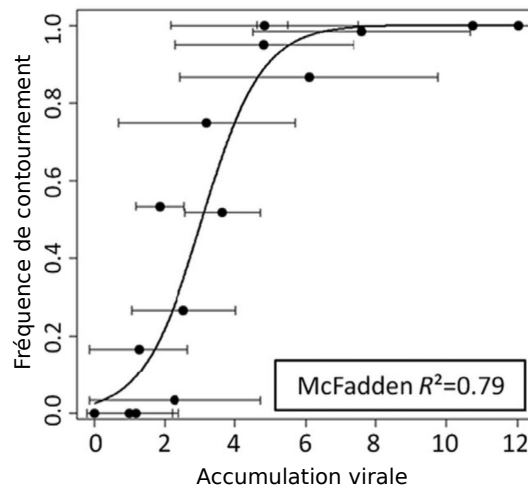


FIGURE 1.10 – Lien entre la fréquence de contournement du gène majeur de résistance *pvr2³* et l'accumulation virale du PVY chez Perennial et 14 lignées HD (issues du croisement entre Perennial et Yolo Wonder). La fréquence de contournement a été mesurée sur 60 plantes par génotype inoculées avec un variant avirulent, et l'accumulation virale a été mesurée sur 10 plantes par génotype inoculées avec un variant virulent. Les points correspondent aux valeurs expérimentales et les barres aux écarts-type. La courbe correspond à un modèle logistique ajusté aux données. Modifié d'après Quenouille et al. (2013).

deux caractères (Fig. 1.9) donne une explication génétique à ce lien significatif. Ainsi, une diminution de l'accumulation virale pourrait conduire à une diminution de la fréquence de contournement de l'allèle majeur de résistance. Ce résultat peut être dû à une diminution du nombre de mutations apparues dans l'ensemble de la population virale lorsque la taille de cette population diminue, ou à une augmentation de la dérive génétique dans les populations de plus petite taille. S'il y a moins de mutations apparues en moyenne dans la population virale, cela signifie une plus faible probabilité d'apparition des mutations de virulence et donc un taux de contournement plus faible. Si l'intensité de la dérive génétique augmente, cela augmente la probabilité d'extinction des éventuelles mutations de virulence apparues dans la population, et donc conduit à un taux de contournement plus faible.

Des voies mutationnelles plus complexes

Une deuxième expérience a consisté à observer les mutations responsables du contournement de *pvr2³* dans le cas d'une résistance monogénique ou polygénique. Pour cela, un variant avirulent a été inoculé par greffage à deux variétés de piment représentatives de chaque situation (monogénique et polygénique). La méthode d'inoculation par greffage consiste à inoculer le virus à un porte-greffe sensible afin qu'il multiplie le virus et produise une source d'inoculum constante pour le greffon (via le système vasculaire) porteur de la résistance étudiée. Le porte-greffe

choisi correspondait à la variété sensible Yolo Wonder et a été inoculé avec le clone avirulent SON41p. L'analyse des populations de PVY pour les greffons de la variété correspondant à la résistance monogénique a conduit à l'identification de quatre variants (G, K, N et C), différant par une substitution nucléotidique non-synonyme du clone initial SON41p dans le gène codant pour la VPg (*virus protein genome-linked*), protéine constituant le facteur d'avirulence vis-à-vis de *pvr2³*. Toutes ces mutations avaient été démontrées comme suffisantes pour contourner la résistance due à *pvr2³* (Ayme et al., 2006). Dans les plantes porteuses de la résistance polygénique, deux variants ont été identifiés, différant chacun par deux substitutions nucléotidiques non-synonymes du clone initial SON41p (GK et KN). Au total, 23 plantes sur 32 inoculées (72%) étaient infectées pour la résistance monogénique six semaines après inoculation, alors que seules 2 plantes sur 91 (2%) inoculées étaient infectées pour la résistance polygénique trois mois après inoculation. Pour résumer, les plantes porteuses de la résistance monogénique ont été contournées par des mutants simples de la VPg, alors que les plantes porteuses de la résistance polygénique ont été contournées par des mutants doubles de la VPg, et la réussite de l'infection était moindre dans les plantes porteuses de la résistance polygénique.

Pour vérifier si deux mutations au lieu d'une sont nécessaires au contournement de la résistance polygénique, trois des mutants simples (G, K et N) ainsi que les deux mutants doubles GK et KN ont été inoculés séparément à des plantes de la variété porteuse de la résistance polygénique, et le pourcentage de plantes présentant une infection systémique a été rapporté pour chaque variant. Les mutants doubles ainsi que le mutant simple K ont présenté des pourcentages d'infection similaires (entre 41 et 46%), alors que les deux autres mutants simples (G et N) ont montré des pourcentages d'infection beaucoup plus faibles (3 et 9%). Quenouille et al. (2013) se sont alors demandés pourquoi le mutant simple K n'avait pas été retrouvé dans les plantes porteuses de la résistance polygénique, puisque son efficacité d'infection est similaire à celles des mutants doubles. Leur hypothèse se base sur le fait que la substitution correspondant à la mutation K est une transversion⁹, contrairement aux deux autres mutants simples G et N pour lesquels la substitution correspondante est une transition¹⁰. Or, une transition est cinq à huit fois plus probable qu'une transversion, ce qui expliquerait pourquoi le mutant simple K n'a pas été observé dans les plantes porteuses de la résistance polygénique. Cela détermine aussi *a priori* le chemin mutational suivi par les virus pour contourner la résistance polygénique. En l'occurrence, ils ont dû d'abord acquérir une des deux substitutions les moins avantageuses mais les plus fréquentes (G ou N), ce qui leur a permis de mieux se multiplier, pour enfin acquérir

9. substitution d'une base purique (A ou G) par une base pyrimidique (C ou T) ou inversement.

10. substitution d'une base purique par une autre base purique, ou d'une base pyrimidique par une autre base pyrimidique.

la deuxième mutation qui était toujours la transversion (K).

Au final, il semblerait donc qu'une résistance quantitative puisse amener à une complexification des voies mutationnelles pour arriver au contournement d'une résistance qualitative, et donc à une moindre fréquence de contournement.

Diminution du taux de fixation du variant adapté à la résistance

Une dernière expérience a consisté à mettre en compétition deux variants, toujours un mutant simple (G, N ou K) avec un mutant double (GK ou KN), dans des plantes porteuses de la résistance monogénique ou polygénique (Fig. 1.11). En tout, quatre couples de variants ont été testés en compétition (GK versus (vs.) G, GK vs. K, KN vs. K et KN vs. N), et les pourcentages de plantes ne montrant que le mutant simple, que le mutant double, ou les deux variants en mélange ont été reportés. Un seul couple sur les quatre (KN vs. K) n'a pas montré de différence de compétitivité entre le mutant simple K, correspondant à une transversion, et le mutant double KN, que ce soit pour la résistance monogénique ou polygénique. Pour les trois autres compétitions, la résistance polygénique entraînait une diminution significative du pourcentage de plantes infectées par le mutant double uniquement, par rapport à la résistance monogénique. Cela signifie que le taux de fixation du variant le plus adapté, c'est-à-dire du mutant double, est moindre dans les plantes porteuses de la résistance polygénique par rapport à celles porteuses de la résistance monogénique. Quenouille et al. (2013) mettent ce résultat en lien avec l'intensité de la sélection et/ou de la dérive génétique. Pour eux, dans les plantes où seul le mutant simple a été retrouvé l'effet de la dérive génétique a été plus fort que celui de la sélection alors que dans les plantes où seul le mutant double a été retrouvé la sélection était la force évolutive majeure agissant sur la population virale. Dans les plantes où les deux variants étaient encore présents, l'intensité des deux forces évolutives devait être plus faible et la fixation d'un des deux variants ne s'était pas encore produite.

Pour résumer, des liens ont été observés entre la fréquence de contournement du gène majeur *pvr2³* et (i) l'accumulation virale, (ii) la complexité des voies mutationnelles impliquées dans le contournement chez le virus, et (iii) la diminution du taux de fixation du variant viral le plus adapté. Cependant, l'importance relative de ces trois facteurs dans la fréquence de contournement n'a pas été estimée.

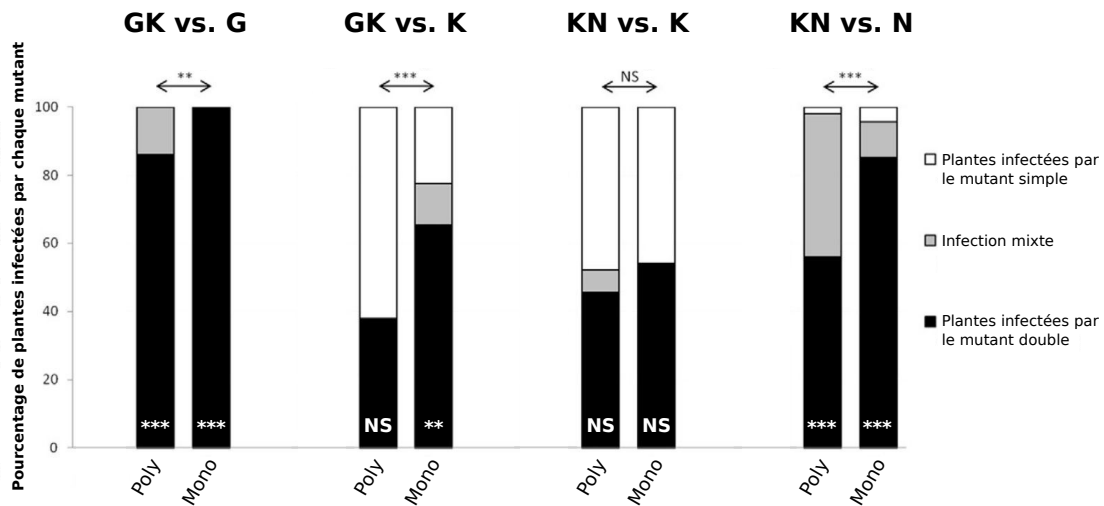


FIGURE 1.11 – Compétitivité de mutants doubles de la VPg (PVY) contre les deux mutants simples correspondants dans des génotypes de piment porteurs d’une résistance polygénique (Poly) ou monogénique (Mono). Les symboles ** et *** dans les barres des histogrammes correspondent à des distributions significativement différentes entre les variants viraux en compétition (mutant simple contre mutant double) dans chaque génotype de plante comparé au ratio 1 :1 de l’inoculum, à un seuil d’erreur de type I de 1% et 0.1% respectivement (test exact de Fisher). Les plantes infectées par les deux variants ont été exclues de cette analyse. Les symboles ** et *** présents au-dessus des histogrammes correspondent à des différences significatives entre génotypes de plante pour les distributions de plantes infectées par le mutant simple, le mutant double ou un mélange des mutants, à un seuil d’erreur de type I de 1% et 0.1% respectivement (test exact de Fisher). NS : pas de différence significative ($p_{value} > 5\%$). Modifié d’après Quenouille et al. (2013).

1.4 Dérive génétique et sélection : des forces évolutives influençant l’adaptation d’un agent pathogène à la résistance des plantes

Au cours de cette introduction, j’ai évoqué les différentes forces évolutives pouvant jouer sur l’évolution d’un agent pathogène – à savoir la mutation, la recombinaison, la sélection, la dérive génétique et la migration – et donné une définition de chacune (encadrés 1.1 à 1.4). Je vais maintenant donner plus de détails sur chacune, notamment pour leur lien avec l’adaptation à un gène majeur de résistance. Je commencerai par les forces évolutives qui permettent de générer la variabilité génétique, à savoir la mutation et la recombinaison. Je continuerai avec les goulets d’étranglement subis par les populations virales dans les plantes, en lien avec la dérive génétique, force évolutive qui tend à diminuer la variabilité génétique. Ensuite, je parlerai de la migration, force évolutive intervenant dans un contexte de métapopulation. Enfin, j’aborderai la sélection et la dérive génétique conjointement, car ces deux forces agissent en interaction et c’est cette interaction qui m’intéresse particulièrement.

1.4.1 Quelques caractéristiques des agents pathogènes en lien avec leur adaptation à résistance des plantes

Mutation

Le taux de mutation varie selon le groupe d'organismes considéré mais aussi au sein d'un même organisme selon la région génomique. Parmi les agents pathogènes des plantes, les eucaryotes (dont font partie les champignons) montrent des taux de mutation par nucléotide et par génération globalement plus faibles (entre 10^{-9} et 10^{-11}) que les bactéries (aux alentours de 10^{-9}), qui se placent elles-même derrière les virus (entre 10^{-4} et 10^{-8}), parmi lesquels les virus à ARN (tel que le PVY) se situent en tête (entre 10^{-4} et 10^{-6}) (Drake and Holland, 1999; Drake et al., 1998; Gago et al., 2009). Enfin, un dernier groupe, constitué par les viroïdes¹¹, montre les taux de mutation les plus élevés, entre 10^{-2} et 10^{-3} (Gago et al., 2009).

Selon le groupe d'organismes considéré, le type de mutation (substitution, insertion ou délétion) principal responsable du contournement d'une résistance ne sera pas le même. Chez les champignons et les bactéries, ce sont majoritairement des insertions ou des délétions, qui entraînent l'inactivation du facteur d'avirulence (Gout et al., 2007; Kang et al., 2001; Kearney et al., 1988; Kousik and Ritchie, 1996; Schürch et al., 2004). Les virus, eux, diffèrent considérablement des agents pathogènes cellulaires de par la nature, la taille et l'expression de leur génome (García-Arenal and McDonald, 2003). Un grand nombre de leurs protéines sont multifonctionnelles, donc une insertion ou une délétion pour l'acquisition de la virulence pourrait avoir des conséquences défavorables sur d'autres fonctions vitales de l'agent pathogène. Ainsi, les substitutions non-synonymes sont le mécanisme principal d'acquisition de la virulence, et dans la plupart des cas un petit nombre de mutations (≤ 3) suffit au contournement de la résistance (Table 1.2) (Harrison, 2002; Janzac et al., 2010; Jenner et al., 2000; Lecoq et al., 2004; Moury et al., 2004).

Recombinaison

McDonald and Linde (2002a) placent la recombinaison comme facteur important de risque pour le contournement des résistances chez les champignons et les bactéries. Elle peut notamment jouer un rôle important lorsque plusieurs gènes de l'agent pathogène sont impliqués dans le contournement d'une résistance, par exemple lorsque plusieurs gènes majeurs de résistance sont pyramidés dans une variété de plante et interagissent avec différents facteurs d'avirulence chez l'agent pathogène (Brown, 2015).

11. Particule plus petite qu'un virus, constituée d'un seul ARN circulaire non codant et sans capsid.

TABLE 1.2 – Déterminants de l'avirulence chez les virus et gènes de résistance correspondant chez les plantes.

Plante	Gène de résistance	Virus	Gène d'avirulence	Changements pour le contournement
<i>Brassica napus</i>	<i>TuRBO1, TuRBO1B</i>	TuMV	CI	1 aa
	<i>TuRBO3</i>	TuMV	P3	1 aa
	<i>TuRBO4, TuRBO5</i>	TuMV	P3	1 aa
<i>Cucumis melo</i> L.	<i>nsv</i>	MNSV	Portion 3' du génome du MNSV	3' UTR
Piment	<i>pvr1¹</i>	PVY	VPg	1 aa
Pomme de terre	<i>Nb</i>	PVX	25 K protéine de mouvement	1 aa
	<i>Nx</i>	PVX	Capside	1 aa
Tabac	<i>Rx1, Rx2</i>	PVX	Capside	1 ou 2 aa
	<i>N'</i>	ToMV	Capside	2 aa
	<i>N</i>	TMV	Réplicase	> 2 aa
Tomate	<i>va</i>	PVY	VPg	1 aa
	<i>Tm-1</i>	ToMV	Réplicase	2 aa
	<i>Tm-2, Tm-2²</i>	ToMV	30kD protéine de mouvement	2 aa
	<i>pot-1</i>	PVY	VPg	1 aa
	<i>Sw-5</i>	TSWV	NSm	> 2 aa

TuMV : *Turnip mosaic virus* (genre *Potyvirus*); MNSV : *Melon necrotic spot virus* (genre *Carmovirus*); PVX : *Potato virus X* (genre *Potyvirus*); ToMV : *Tomato mosaic virus* (genre *Tobamovirus*); PVY : *Potato virus Y* (genre *Potyvirus*); TSWV : *Tomato spotted wilt virus* (genre *Tospovirus*); UTR : région non traduite (*untranslated region*); aa : acide aminé. Modifié d'après Kang et al. (2005)

La co-infection d'une même cellule par au moins deux variants différents est nécessaire pour qu'il y ait recombinaison chez les virus de plante. La fréquence de ces co-infections est très variable d'un virus à un autre ; par exemple une ségrégation spatiale forte a été observée chez le *Turnip mosaic virus* (genre *Potyvirus*) (TuMV) alors que 100% des cellules étaient co-infectées en systémie chez le *Cauliflower mosaic virus* (genre *Caulimovirus*) (CaMV) (Gutiérrez et al., 2010, 2012a). À ma connaissance, la recombinaison comme mécanisme de contournement d'une résistance par des virus de plantes n'a été rapporté que deux fois dans la littérature. Le premier cas concerne l'adaptation du *Melon necrotic spot virus* (genre *Carmovirus*) (MNSV) au gène de résistance *nsv* du melon (*Cucumis melo*), dont la recombinaison nécessaire au contournement concerne la région 3' non traduite (UTR, *untranslated region*) du génome viral (Table 1.2) (Díaz et al., 2004; Miras et al., 2014). Le deuxième cas concerne l'émergence de recombinants de PVY, qui pourrait être liée à l'utilisation de variétés de pomme de terre porteuses des gènes de résistance *Nc* et/ou *Ny* (Moury et al., 2011a). Par ailleurs, deux études ont rapporté des taux de recombinaison élevés chez le CaMV, de l'ordre de $2 \cdot 10^{-5}$ à $4 \cdot 10^{-5}$ par base et par cycle de réplication, et de l'ordre de $3.4 \cdot 10^{-5}$ par base et par cycle de réplication dans les cellules co-infectées par plusieurs variants chez le *Tobacco etch virus* (genre *Potyvirus*) (TEV), indiquant des recombinaisons courantes chez ces virus qui pourraient conduire à un risque de contournement plus élevé (Froissart et al., 2005; Tromas et al., 2014).

Dérive génétique et goulets d'étranglement chez les virus de plante

Les goulets d'étranglement correspondent à des diminutions de tailles de populations, et par conséquent également de tailles efficaces de population (Encadré 1.3). Ainsi, ils augmentent l'intensité de la dérive génétique agissant sur les populations d'agents pathogènes, ce qui tend à diminuer leur diversité génétique. En conséquence, les goulets d'étranglement pourraient constituer un frein à l'adaptation des populations d'agents pathogènes à un gène majeur de résistance (McDonald and Linde, 2002a). McDonald and Linde (2002a) et García-Arenal and McDonald (2003) placent la taille efficace de population comme facteur de risque pour le potentiel évolutif des agents pathogènes face à la résistance des plantes. Ces auteurs préconisent des stratégies de contrôle gardant les tailles efficaces de population d'agents pathogènes petites pour limiter leur diversité génétique.

Les virus de plantes sont tout particulièrement concernés par les goulets d'étranglement, qu'ils subissent à l'intérieur de leurs plantes hôtes. En effet, l'infection d'une plante par un virus est un cheminement complexe pendant lequel des goulets d'étranglement peuvent survenir lors

de différentes étapes (Elena et al., 2008; Gutiérrez et al., 2012a). À l'inoculation, des estimations de tailles efficaces de population très faibles suite à la transmission de virus par des pucerons vecteurs ont été reportées, variant de 0.5 à 2.2 en moyenne (pathosystèmes PVY - piment et *Cucumber mosaic virus* (genre *Cucumovirus*) (CMV) - tomate) (Betancourt et al., 2008; Moury et al., 2007). Durant la colonisation des feuilles par mouvement de cellule à cellule, le nombre d'unités de génome viral infectant une cellule, autrement connu sous le nom de multiplicité d'infection cellulaire (*multiplicity of cellular infection*, MOI), a été également estimé à des valeurs très faibles, variant entre 1 et 13 pour les deux pathosystèmes TMV - tabac (*Nicotiana benthamiana*) et CaMV - navet (*Brassica rapa*) (González-Jara et al., 2009; Gutiérrez et al., 2010). Durant l'infection systémique, i.e. la colonisation de toutes les feuilles de la plante via le système vasculaire, de forts goulets d'étranglement ont aussi été reportés, les tailles efficaces de population variant entre 1 et 5.6 pour les pathosystèmes *Wheat streak mosaic virus* (genre *Tritimovirus*) (WSMV) - blé (*Triticum* spp.) (French and Stenger, 2003), TMV - tabac (Sacristán et al., 2003), PVY - piment (Fabre et al., 2012a) et TEV - piment (Zwart et al., 2011). Une estimation beaucoup plus élevée a été publiée chez le pathosystème CaMV - navet (298 - 484) (Monsion et al., 2008), mais celle-ci a ensuite été ré-évaluée à la baisse (8.8 - 131) (Gutiérrez et al., 2012b). Une dernière estimation intermédiaire a été reportée, entre 1.2 et 47.9, pour le couple TEV - tabac (Zwart et al., 2011).

Les goulets d'étranglement sont donc nombreux chez les virus de plante, lors de la transmission inter-hôtes mais aussi lors de la colonisation des hôtes, lors du mouvement de cellule à cellule ou du passage d'une feuille à une autre (infection systémique). Ils peuvent être extrêmes et jouer un rôle important pour l'évolution des populations virales (Zwart and Elena, 2015). Comme je l'ai déjà dit, les goulets d'étranglement ont une influence directe sur l'intensité de la dérive génétique agissant sur les populations d'agents pathogènes. Nous verrons un peu plus loin l'effet d'une augmentation de la dérive génétique sur le potentiel adaptatif des agents pathogènes aux gènes majeurs de résistance des plantes.

Par ailleurs, au niveau de l'agroécosystème la saisonnalité des cultures et/ou des conditions climatiques difficiles ont pour conséquence la présence de goulets d'étranglement chez tous les agents pathogènes de plante concernés. Des goulets d'étranglement peuvent donc être présent à différents niveaux, en particulier chez les virus de plantes, depuis le niveau intra-hôte jusqu'au niveau d'une population d'hôtes. Les tailles efficaces de populations qui en découlent sont importantes aux deux niveaux. En effet, les dynamiques de populations d'agents pathogènes dans les hôtes comme leurs dynamiques épidémiques sont déterminantes pour l'émergence de variants pathogènes adaptés à des résistances des plantes (Galvani, 2003; Mideo et al., 2008).

Migration

La force évolutive de migration intervient dans un contexte de métapopulation¹². Les échelles auxquelles les agents pathogènes se dispersent sont très variables selon les organismes, mais aussi selon l'intervention de l'homme, qui accroît les distances de dispersion au delà de leurs limites naturelles, via les pratiques agricoles et le commerce (McDonald and Linde, 2002a). Ces variabilités et modifications ont des conséquences sur les dynamiques épidémiques, sur la persistance de l'agent pathogène, ainsi que sur la co-évolution entre l'hôte et son agent pathogène (Thrall and Burdon, 1999), et jouent donc un rôle sur l'adaptation d'un agent pathogène à un gène de résistance chez les plantes. Notamment, McDonald and Linde (2002a) et Brown (2015) ont avancé que les agents pathogènes montrant des migrations à longue distance, tels que les champignons de la rouille et de l'oïdium, présentaient un plus grand risque pour le contournement des résistances que ceux montrant des migrations à faible distance.

Lenormand (2002) s'est attaché à étudier l'effet de la sélection et de la migration sur l'adaptation à un environnement, contexte pouvant typiquement être transposé à l'adaptation d'un agent pathogène à une résistance chez la plante. Cet auteur a montré que, dans un environnement hétérogène, la migration depuis les autres habitats s'opposait à l'adaptation locale et conduisait à une diminution de la fitness moyenne des populations locales (Lenormand, 2002; Papaix et al., 2013). De manière intéressante, deux autre études – dont une s'appuyait sur une expérience biologique – ont montré à l'inverse que la migration dans une métapopulation favoriserait l'adaptation locale, tant que cette migration n'homogénéise pas complètement les populations (Gandon and Michalakis, 2002; Morgan et al., 2005). Enfin, Whitlock and Gomulkiewicz (2005) ont étudié la probabilité de fixation de mutations adaptatives dans une métapopulation où la sélection est hétérogène entre habitats. Ils ont conclu que les taux de migration les plus faibles, respectivement les plus forts, conduisaient à aux probabilités de fixation les plus faibles, respectivement les plus fortes (Patwa and Wahl, 2008; Whitlock and Gomulkiewicz, 2005).

1.4.2 Interaction entre dérive génétique et sélection

Je vais maintenant décrire plus en détails la sélection et la dérive génétique, qui sont les deux forces évolutives qui m'ont particulièrement intéressée pendant ma thèse. Je vais ici les aborder en même temps, car ces deux forces agissent de manière conjointe sur l'évolution d'une

12. Ensemble de populations d'une même espèce réparties dans l'espace et connectées entre elles, ces connections correspondant à des migrations d'individus entre populations.

population. Nous verrons qu'à cause de cette caractéristique il est difficile de les dissocier et de savoir quels sont les rôles relatifs de la dérive génétique et de la sélection dans l'évolution d'une population (Foll et al., 2014c).

Le modèle de Wright-Fisher

En règle générale, la dérive génétique s'oppose à la sélection. L'efficacité de la sélection pour la propagation d'une mutation favorable – c'est-à-dire sous sélection positive –, ou pour l'élimination d'une mutation délétère – c'est-à-dire sous sélection négative – est contrôlée par le produit $N_e s$ entre la taille efficace de population N_e et le coefficient de sélection s (Charlesworth, 2009). Cela peut être illustré grâce au célèbre modèle de Wright et Fisher, fondateurs de la génétique des populations (Fisher, 1922, 1930; Wright, 1931). Ce modèle suppose que l'évolution du nombre d'individus avantagés dans une population sous l'effet de la dérive génétique et de la sélection suit une loi binomiale, telle que :

$$X_{n+1} = \mathcal{B} \left(N_e, \frac{(1+s)X_n}{(1+s)X_n + N_e - X_n} \right), \quad (1.1)$$

où X_n est le nombre d'individus avantagés à la génération n et s est le coefficient de sélection du variant avantagé. À chaque nouvelle génération, la composition de la population est déterminée par un tirage aléatoire avec remise des descendants (des gamètes pour les eucaryotes) produits par la génération précédente. Le nombre d'individus se reproduisant (ou se multipliant pour les virus) et contribuant à la génération suivante est donné par N_e . À cause de la dérive génétique, dont l'intensité est reliée à la valeur de N_e , la fréquence d'un allèle peut aléatoirement augmenter ou diminuer d'une génération à la suivante (Malaspinas, 2016).

En guise d'illustration, j'ai réalisé des simulations du modèle de Wright-Fisher pour trois régimes de sélection et de dérive génétique différents, dans un contexte où le variant avantagé est initialement présent à une faible fréquence (1%) pour mimer des scénarios d'évolution de populations d'agents pathogènes suite à l'introduction d'une variété résistante (Fig. 1.12). Dans le premier cas, la sélection est forte ($s = 0.1$) et la dérive génétique est faible ($N_e = 1000$). Les réplicats de simulations pour cette condition montrent une majorité de cas où la fréquence du variant avantagé augmente au cours des générations, jusqu'à avoir complètement envahi la population dans la majorité des cas au bout de 100 générations (pour plus de 80% des simulations). Dans seulement environ 10% des simulations, le variant adapté disparaît. Dans le second cas, la sélection est dix fois plus faible ($s = 0.01$) et la taille efficace de population est la même. Cette fois-ci la tendance s'inverse, aucun des 100 réplicats de simulations ne voit

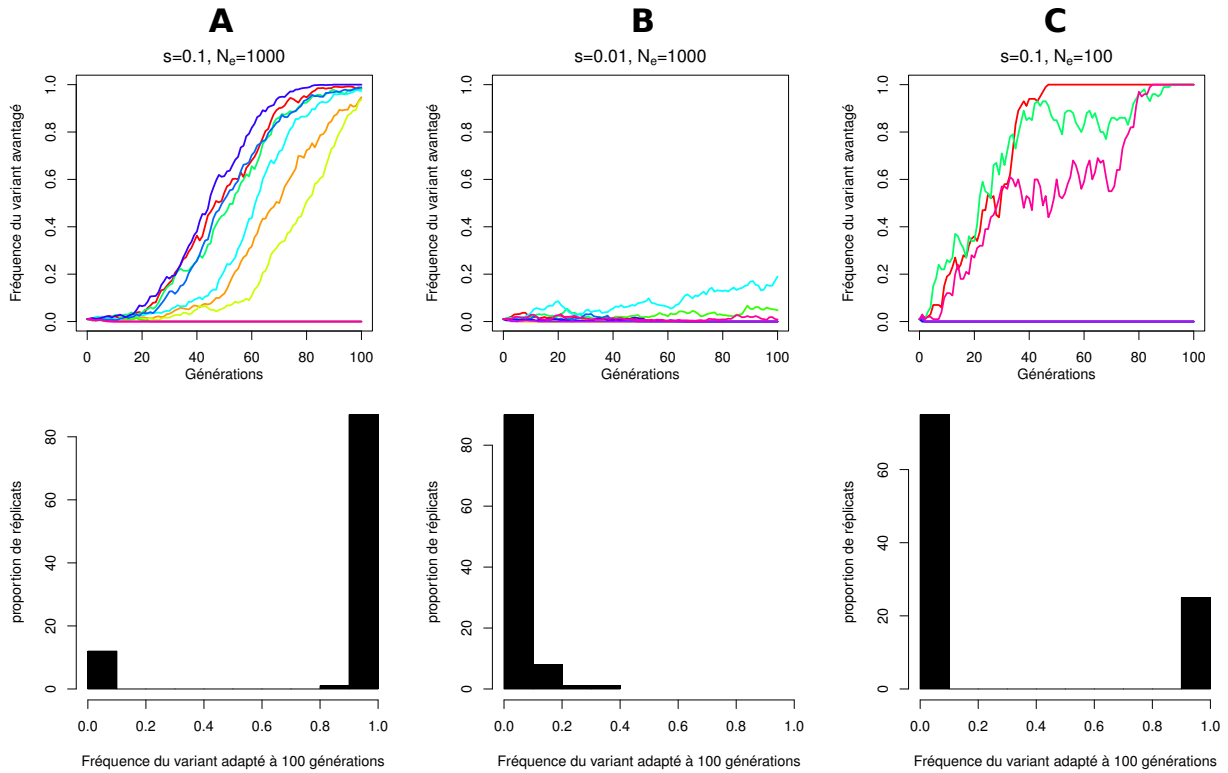


FIGURE 1.12 – Simulations d'évolutions de la fréquence d'un allèle avantageux au cours de 100 générations selon le modèle de Wright-Fisher (première ligne de graphiques) et distribution des fréquences de cet allèle adapté au bout des 100 générations sur 100 réplicats de simulations (deuxième ligne). Trois régimes de sélection et de dérive génétique sont représentés. **A** : forte sélection, faible dérive génétique ; **B** : faible sélection, faible dérive génétique ; **C** : forte sélection, forte dérive génétique. Pour les graphiques d'évolution, seuls 10 réplicats parmi 100 ont été représentés, tirés au hasard. Les valeurs du coefficient de sélection s et de la taille efficace de population N_e figurent en haut de chaque graphique d'évolution. Toutes les simulations sont initiées avec 1% d'individus avantageux.

le variant adapté envahir la population. Au mieux, le variant adapté arrive à une fréquence de 0.2 - 0.3 au bout de 100 générations. Dans la majorité des cas (environ 90%), le variant adapté disparaît. Nous pouvons maintenant observer l'effet d'une augmentation de la dérive génétique, pour le troisième cas, avec une forte sélection ($s = 0.1$) et une forte dérive génétique ($N_e = 100$). Les courbes sont plus chaotiques (en dents de scie) à cause de la dérive génétique. L'effet de la dérive génétique sur la variabilité génétique d'une population est bien visible : au bout des 100 générations, le variant adapté a soit complètement envahi la population (dans environ 25% des cas) soit complètement disparu (dans environ 75% des cas).

En résumé, le variant avantageux peut disparaître par des effets de stochasticité due à la dérive génétique, et s'il ne disparaît pas, il suivra une trajectoire moyenne avec une amplitude d'oscillations qui dépend de la valeur de N_e . Plus N_e est petit et plus les variations de fréquence d'une génération à la suivante pourront être grandes et les oscillations importantes. Plus N_e est grand, plus l'effet de la dérive génétique est faible et plus les trajectoires sont similaires

entre elles (Malaspinas, 2016). Selon le modèle de Wright-Fisher dans le cas de sélection neutre, la variance des variations de fréquence d'une génération 0 à la suivante est inversement proportionnelle à N_e , et vaut p_0q_0/N_e , avec p_0 la fréquence du variant avantageé à la génération 0 et q_0 celle du variant désavantageé ($q_0 = 1 - p_0$) (Charlesworth, 2009).

La figure 1.12 illustre bien la difficulté de dissocier les effets de la sélection et de la dérive génétique à partir de telles trajectoires évolutives. Même si les deux forces agissent différemment, la sélection positive tendant à faire augmenter la fréquence du variant adapté et la dérive génétique conduisant à des fluctuations dans les fréquences, les différents types de trajectoires sont souvent difficiles à distinguer (Malaspinas, 2016). Dans ce contexte, Rouzine et al. (2001) ont défini différents régimes de sélection et de dérive génétique auxquels ils ont associé différents comportements types. La force évolutive de mutation est incluse dans leur analyse. Je vais décrire succinctement ces régimes pour des cas incluant la sélection (le coefficient de sélection est supérieur au taux de mutation). Selon ces auteurs, si $N_e \ll 1/s$, alors les trajectoires des fréquences sont stochastiques et la population suit un régime de dérive génétique. Dans ce cas, la sélection peut être négligée. À l'opposé, si N_e est très grand (N_e supérieur au taux de mutation), alors les trajectoires sont déterministes et la population suit un régime de sélection ; la dérive génétique peut être négligée. Si N_e est intermédiaire entre ces deux extrêmes, les trois forces évolutives (mutation, sélection, dérive génétique) jouent un rôle significatif sur l'évolution de la population. La mutation crée de nouveaux variants, la sélection agit sur ces variants, et la dérive génétique peut produire de fortes fluctuations dans, et entre les trajectoires.

Probabilité de fixation de mutations bénéfiques

J'ai évoqué à plusieurs reprises la notion de taux de fixation ou de probabilité de fixation de mutations. De nombreuses études théoriques s'attachent à l'étude de ce concept et à en définir des expressions. Les histogrammes de la figure 1.12, qui représentent la distribution de la fréquence d'un allèle avantageé au bout d'un certain nombre de générations, peuvent être considérés comme des estimations du taux de fixation de cet allèle (proportion des simulations qui aboutissent à une fréquence de 1). D'ailleurs, pour faire le parallèle avec les interprétations de Quenouille et al. (2013), il est clairement visible qu'une diminution du coefficient de sélection ou une diminution de la taille efficace de population conduisent à un taux de fixation plus faible, donc une sélection ralentie.

Pour formaliser cette notion de probabilité de fixation, je vais présenter la célèbre expression de Kimura (1955, 1962, 1970), valable dans le cas d'une population haploïde à reproduction

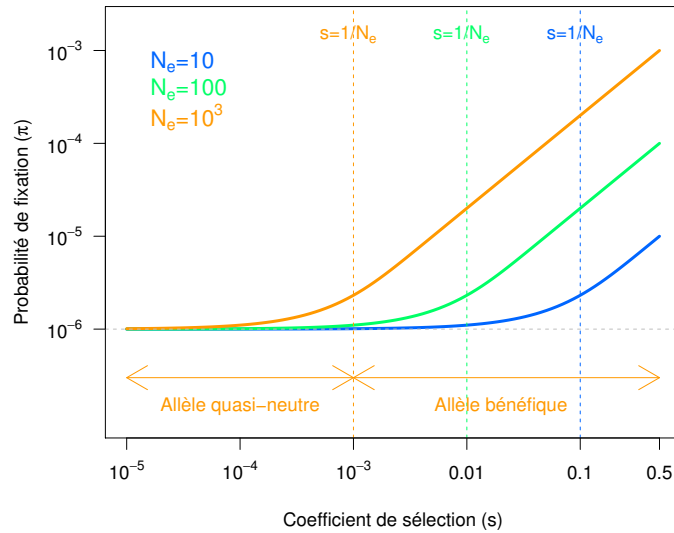


FIGURE 1.13 – Probabilité de fixation (π) d'un allèle favorable dans une population haploïde pour différents régimes de sélection (coefficient de sélection s) et de dérive génétique (taille efficace de population N_e). La fréquence initiale de l'allèle favorable est toujours fixée à $1/N$, avec la taille de population $N = 10^6$, soit une mutation apparue en une copie. Trois régimes de dérive génétique sont testés grâce à trois valeurs de taille efficace de population : $N_e = 10$, 100 et 10^3 . Les droites en pointillés verticales sont situées en $1/N_e$, délimitant la valeur du coefficient de sélection pour laquelle on passe d'un allèle considéré comme quasi-neutre à un allèle considéré comme bénéfique (plus forte probabilité de fixation). La droite horizontale en pointillés gris correspond à la probabilité de fixation d'un allèle neutre, $1/N$.

sexuée, panmictique, de grande taille constante, à générations discrètes, avec un nombre de descendants poissonien et considérant la présence d'un seul mutant à la fois dans la population (mutation rare, pas plus de deux variants présents simultanément dans la population). L'hypothèse du nombre de descendants poissonien signifie que le nombre de descendants fournis par chaque individu pour la génération suivante suit une loi de Poisson. Cette hypothèse en cache une autre : les individus ne peuvent pas mourir avant de produire une descendance (Patwa and Wahl, 2008). Kimura (1955, 1962, 1970) fournit une approximation de la probabilité de fixation π d'un allèle favorable de coefficient de sélection s qui se trouve initialement à une fréquence p dans une telle population :

$$\pi = \frac{1 - e^{-2spN_e}}{1 - e^{-2sN_e}}. \quad (1.2)$$

Cette équation se base sur une approximation de la diffusion, ce qui implique qu'elle est valide pour des grandes tailles de population et des petits changements de fréquence allélique à chaque génération, soit une faible sélection (Patwa and Wahl, 2008). J'ai réalisé un graphique de la probabilité de fixation π pour un continuum de valeurs de s et différentes valeurs de N_e (Fig. 1.13). Je me suis placée dans le cas où l'allèle favorable apparaît en une copie dans la population, par exemple suite à une mutation. Ainsi, p a été fixé à $1/N$ ($N = 10^6$), avec N la taille réelle

de la population. À partir d'un seuil situé à environ $s = 1/N_e$, la probabilité de fixation augmente avec le coefficient de sélection s . Ce seuil dépend de la valeur de N_e : plus N_e est petit et plus le seuil se situe vers des plus grandes valeurs de s . En-dessous du seuil, l'allèle favorable étudié devient quasi-neutre, ce qui signifie que la sélection devient négligeable (régime de dérive génétique défini par Rouzine et al. (2001)). Au-dessus du seuil, la sélection n'est plus négligeable, on tombe dans le régime sélection - dérive génétique de Rouzine et al. (2001). Cela rejoint également les remarques évoquées dans la partie précédente sur l'importance du produit $N_e s$ pour déterminer l'issue de l'évolution (Charlesworth, 2009). De plus, une diminution de N_e réduit la probabilité de fixation de l'allèle favorable (Patwa and Wahl, 2008). Cet effet est d'autant plus visible dans des gammes de valeurs de s pour lesquelles la sélection n'est pas négligeable (régime sélection - dérive génétique). Pour de très faibles valeurs de s ($s < 1/N_e$), toutes les probabilités de fixation convergent vers une même valeur, $1/N$, qui correspond à la probabilité de fixation d'un allèle neutre.

Lorsque les mutations ne sont pas rares et que plus de deux mutants peuvent être présents simultanément dans la population, on parle d'interférence clonale. Muller (1932, 1964) a prédit que l'interférence clonale dans les populations asexuées, conduisant à une compétition entre plusieurs mutants avantageés, tendait en général à diminuer π , et qu'il existait une vitesse limite au taux d'adaptation dans de telles populations. Plus tard, Gerrish and Lenski (1998) ont conclu qu'une mutation bénéfique arrive à fixation si et seulement si (i) elle survit à la dérive génétique et (ii) qu'aucune mutation à effet de sélection supérieur n'apparaît et ne survit à la dérive génétique pendant l'intervalle de temps requis pour sa fixation. Leur théorie se base sur l'hypothèse de mutants simples du variant sauvage uniquement, i.e. ils ne considèrent pas l'effet de variants accumulant plus d'une mutation par rapport au variant sauvage. Sniegowski and Gerrish (2010) suggèrent que lorsqu'une mutation bénéfique devient commune dans une population, il devient plus probable que des variants portants de multiples mutations bénéfiques apparaissent. Ces auteurs ont développé un modèle pour tenir compte de cette possibilité de mutants multiples, et ont montré que des nuages de mutants se fixaient dans la population, par vagues isolées les unes des autres.

Goulets d'étranglement et populations de tailles variables

Dans des populations dont la taille efficace de population varie dans le temps, Otto and Whitlock (1997) ont montré que la formule de Kimura (1955, 1962) s'appliquait à chaque génération, en remplaçant N_e par la moyenne harmonique des tailles efficaces de population

depuis la première génération jusqu'à la génération étudiée, sous des conditions de coefficient de sélection petit et de grande taille efficace de population. Les variations des tailles de populations qu'ils ont étudié suivaient une équation logistique (la taille de la population ne peut pas dépasser un seuil correspondant à sa capacité limite). Otto and Whitlock (1997) ont fourni une expression de N_e pour laquelle la formule de Kimura (1955, 1962) restait valable pour de grands coefficients de sélection et de petites tailles efficaces de population.

Wahl and Gerrish (2001) et Wahl et al. (2002) ont étudié l'effet de goulets d'étranglement sur la probabilité de fixation d'une mutation bénéfique, avec des périodes de croissance soit exponentielle, soit logistique, entre deux goulets d'étranglement. Ces deux études ont conduit à des conclusions similaires, à savoir que plus la mutation survenait tardivement après un goulet d'étranglement, moins elle avait de chance de survivre aux prochains goulets d'étranglement. Enfin, Heffernan and Wahl (2002) ont ajouté l'effet de la dérive génétique au début des phases de croissance après chaque goulet d'étranglement, et ont démontré que la probabilité de fixation était alors réduite d'au moins 25% par rapport aux résultats de Wahl and Gerrish (2001) et Wahl et al. (2002).

On voit bien l'intérêt de ces résultats par rapport aux nombreux goulets d'étranglement subis par les virus dans leurs plantes hôtes. Cependant, nous avons aussi vu que dans le cas d'une résistance qualitative suivant le modèle gène-pour-gène, le coefficient de sélection entre un variant avirulent et virulent dans une plante résistante devrait théoriquement être infini, et en pratique très élevé, ce qui constituerait un cas limite des résultats des études que je viens d'évoquer. Par contre, comme nous l'avons vu pour les résultats de Quenouille et al. (2013, 2014) sur les résistances polygéniques, les mutants simples G et N sont moins bien adaptés à ces résistances que les mutants doubles GK et KN, et les mutants simples constituent *a priori* la première étape vers l'acquisition de la deuxième mutation K. Ainsi, le cas de la transition entre mutant simple et mutant double, pour lequel le coefficient de sélection entre ces deux types de variants est moins élevé qu'entre un variant avirulent et virulent, doit être plus approprié aux résultats sur les probabilités de fixation de mutations bénéfiques que je viens de décrire.

Front de science : limites actuelles des modèles d'estimation de N_e et s

L'estimation conjointe de N_e et s dans une population est une thématique d'actualité qui n'est pas évidente à traiter. Nous pouvons maintenant obtenir des données d'évolution moléculaire à un coût de plus en plus réduit grâce aux avancées techniques dans le séquençage haut débit (Lanfear et al., 2014). Ces avancées voient naître des modèles mathématiques qui

ont pour but d'estimer des tailles efficaces de population et/ou des coefficients de sélection (Table 1.3). Comme tous les modèles, ceux-ci sont basés sur des hypothèses, et parfois ces hypothèses ne peuvent pas être validées pour tous les modèles biologiques étudiés. Notamment, les modèles basés sur une approximation de la diffusion supposent que la taille efficace de population est grande et le coefficient de sélection petit (effets faible de la sélection et de la dérive génétique) (Bollback et al., 2008; Malaspinas et al., 2012; Steinrücken et al., 2014). Ces deux hypothèses ne sont pas toujours vérifiées, particulièrement lorsque l'on étudie l'adaptation des agents pathogènes à une résistance qualitative, situation dans laquelle les coefficients de sélection entre variants peuvent être très grands, tel qu'évoqué précédemment. Aussi, lorsque l'on étudie les virus de plantes, lesquels les goulets d'étranglements peuvent être très étroits dans les plantes. Le modèle de Foll et al. (2014b,c) permet de relâcher ces hypothèses car il ne se base pas sur une approximation de la diffusion, mais il requiert néanmoins qu'au moins 90% des marqueurs analysés soient sélectivement neutres. Là encore, cette hypothèse paraît difficile à valider dans le cas des virus de plante, qui présentent des génomes de petite taille et peu de sites sélectivement neutres (Carrasco et al., 2007; Elena et al., 2014; Sanjuán et al., 2004). Enfin, le modèle proposé par Lacerda and Seoighe (2014) suppose que la taille efficace de population ne soit pas trop petite. Typiquement, dans leur étude, elle n'est jamais inférieure à 1000. Les tailles de goulets d'étranglement chez les virus de plantes vont bien en-deça de cette limite. Ainsi, les études de co-estimation de tailles efficaces de population et de coefficients de sélection sont en plein essor, mais il faut encore développer de nouvelles méthodes qui pourront s'appliquer pour des organismes qui ne respectent pas les hypothèses requises jusqu'à présent.

TABLE 1.3 – Modèles estimant taille efficace de population N_e et coefficient de sélection s .

Référence	Modèle et approximation	Quelques hypothèses	Un ou plusieurs loci	Paramètres estimés ^a
Bollback et al. (2008)	WF, approx. diffusion	N_e grand, s petit	Un	$2N_e s$
Malaspinas et al. (2012)	WF, approx. diffusion	N_e grand, s petit	Un	s
Mathieson and McVean (2013)	WF, approx. Gaussienne	s petit, $0 < x_i < 1$ ^b	Un	s
Foll et al. (2014b,c)	WF	$\geq 90\%$ marqueurs neutres	Plusieurs	N_e, s
Lacerda and Seoighe (2014)	WF, méthode delta	N_e grand	Un	N_e, s
Steinrücken et al. (2014)	WF, approx. diffusion	N_e grand, s petit	Un	N_e, s
Terhorst et al. (2015)	WF, approx. Gaussienne	$0 < x_i < 1$	Plusieurs	soit N_e , soit s

WF : Wright-Fisher.

x_i : fréquence d'un allèle i dans la population étudiée.

^a : seules les estimations des paramètres N_e et s sont indiquées ici.

^b : les fréquences des allèles sont supposées rester loin des bornes d'extinction ou de fixation.

Modifié d'après Malaspinas (2016).

1.4.3 Hypothèses sur le lien entre la fréquence de contournement du gène majeur de résistance *pvr2*³ du piment et les forces évolutives agissant sur le PVY

Après avoir donné des détails sur les forces évolutives chez les agents pathogènes des plantes et dans la théorie plus générale de la génétique des populations, je reviens ici sur les résultats de Quenouille et al. (2013) sur le pathosystème piment - PVY, présentés dans la partie 1.3.2. Pour rappel, Quenouille et al. (2013) ont montré que les résistances quantitatives présentes dans le fond génétique de plantes porteuses d'une résistance qualitative permettaient de (i) réduire l'accumulation du virus dans les plantes, (ii) complexifier les voies mutationnelles pour le contournement de la résistance qualitative et (iii) diminuer le taux de fixation du variant le plus adapté à la résistance qualitative. Quenouille et al. (2013) ont proposé que trois mécanismes étaient impliqués dans la durabilité renforcée de la résistance polygénique (Fig. 1.14).

Quenouille et al. (2013) ont relié chacun des trois mécanismes validés aux forces évolutives exercées sur les populations de PVY (Fig. 1.14). Dans le cas de la résistance polygénique, la plus grande efficacité de la résistance fait référence au lien positif entre l'accumulation virale et la fréquence de contournement, la modification du déterminisme génétique du contournement à la complexification des voies mutationnelles, et la diminution de la sélection à la diminution du taux de fixation du variant le plus adapté. La plus grande efficacité de la résistance est due à des fonds génétiques qui renforcent le niveau de résistance et qui diminuent la multiplication de l'agent pathogène (diminution de l'accumulation virale). Cela entraînerait une diminution de la probabilité d'apparition de mutations, mais aussi potentiellement une diminution de la taille efficace de population de l'agent pathogène. Le déterminisme génétique du contournement requiert un nombre plus important de mutations chez le virus, notamment des transversions qui ont moins de chances d'apparaître à cause de la plus faible multiplication du virus. Tous ces processus conduisent à des chemins mutationnels pour le contournement moins nombreux ou plus complexes. La diminution du coefficient de sélection entre les variants plus ou moins adaptés et/ou l'augmentation de la dérive génétique (N_e plus faibles) pourraient conduire à une sélection plus lente des variants viraux les plus adaptés. Au final, l'ensemble de ces processus participerait à une augmentation de la durabilité de la résistance qualitative.

Les hypothèses qui m'intéressent particulièrement sont celles d'une diminution du coefficient de sélection et/ou d'une augmentation de la dérive génétique dans le cas de la résistance polygénique. Il est important d'évaluer les contributions relatives de ces deux forces évolutives

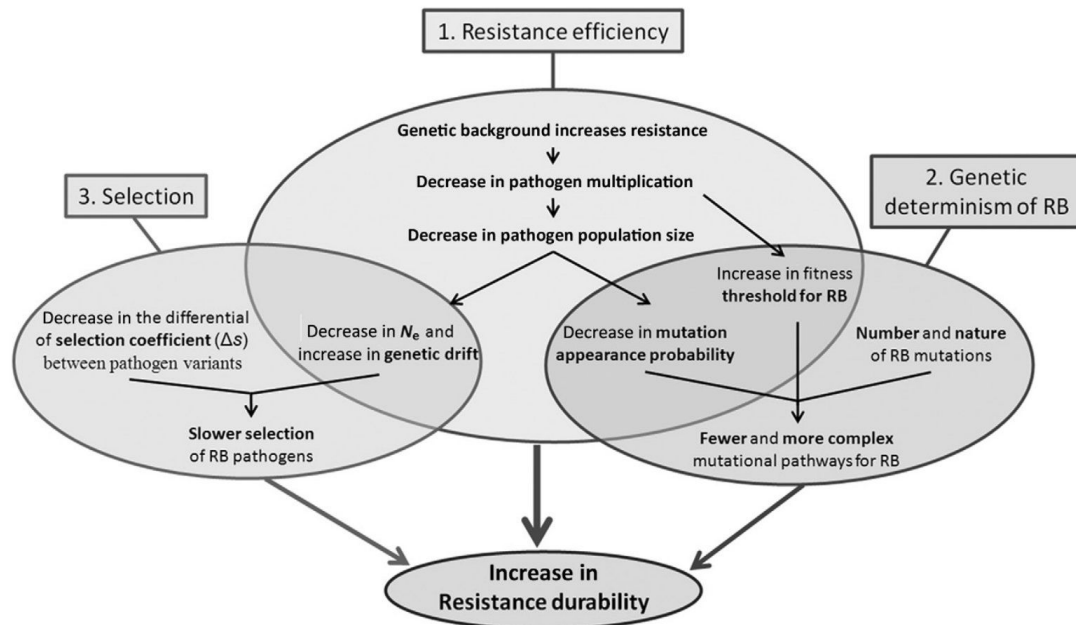


FIGURE 1.14 – Effet de résistances quantitatives conférées par le fond génétique sur la durabilité d'un gène majeur de résistance. Le schéma représente l'ensemble des liens hypothétiques entre les trois mécanismes validés pour leur implication dans la plus forte durabilité d'une résistance polygénique par rapport à une résistance monogénique. N_e : taille efficace de population, RB : contournement de la résistance qualitative, Δs : différence de coefficients de sélection entre les variants *vir* et *avr*. Source : (Quenouille et al., 2013).

sur le ralentissement de la fixation des agents pathogènes les plus adaptés pour pouvoir ensuite étudier leur impact à l'échelle de l'agroécosystème et proposer leur exploitation en amélioration variétale.

1.4.4 Applications à l'échelle d'un agro-écosystème

L'intégration de connaissances sur les principes éco-évolutifs guidant l'interaction entre un hôte et son pathogène est une piste avancée de plus en plus souvent pour une meilleure gestion des maladies dans les cultures (Brown, 2015; García-Arenal et al., 2001, 2003; McDonald and Linde, 2002a; Zhan et al., 2015). L'importance de la dérive génétique dans les processus évolutifs appliqués aux interactions de type gène-pour-gène a été discutée depuis de nombreuses années (Thompson and Burdon, 1992). Notamment, les études correspondantes ont suggéré que la dérive génétique et les flux de gènes entre populations d'agents pathogènes pouvaient être au moins aussi importants que la sélection locale pour les dynamiques génétiques des agents pathogènes et leur vitesse d'adaptation à la résistance des plantes (Thompson and Burdon, 1992). Comme évoqué précédemment, McDonald and Linde (2002a) et García-Arenal and McDonald (2003) avancent que des stratégies de contrôle gardant les tailles efficaces de

population des agents pathogènes petites permettraient de diminuer leur potentiel évolutif face à la résistance des plantes. Récemment, Zhan et al. (2015) ont rediscuté de l'intérêt d'augmenter l'intensité de la dérive génétique pour mieux préserver une résistance, et ont proposé plusieurs solutions pour y arriver : de longues périodes de jachère entre les saisons de culture, des pratiques prophylactiques dans les parcelles (telles que les pratiques du brûlis ou de l'enterrement des résidus de cultures ; l'élimination des réservoirs ou des plantes infectées en cours de culture), les mélanges de variétés et la rotation des cultures (Fig. 1.5) (Zhan et al., 2014). Dans le cadre de cette thèse, je propose une autre solution, qui consiste à se concentrer sur le fond génétique de l'hôte pour jouer sur un trait quantitatif directement relié à l'intensité de la dérive génétique, à savoir la taille efficace de population de l'agent pathogène.

Pour cela, j'ai étudié l'effet de la dérive génétique et de la sélection sur l'émergence de variants virulents et sur la durabilité d'une résistance qualitative depuis l'échelle intra-plante jusqu'à l'échelle d'une parcelle cultivée. Dans la suite de ce manuscrit, je présenterai trois articles qui m'ont permis d'aborder mon sujet de thèse aux différentes échelles étudiées. Dans un premier temps, j'ai développé un modèle épidémiologique stochastique à l'échelle de la parcelle pour étudier et prédire l'effet de résistances quantitatives contrôlant les tailles efficaces de populations virales dans le génome des plantes sur la durabilité d'une résistance qualitative (article 1). Dans un deuxième temps, j'ai étudié les dynamiques de variants viraux en compétition dans des plantes présentant des fonds génétiques variés mais portant la même résistance qualitative (article 2). J'ai développé et validé un modèle mathématique permettant d'estimer les coefficients de sélection et les tailles efficaces de population sur ce type de données. Dans un troisième temps, j'ai effectué une analyse globale des données de Quenouille et al. (2014) de fréquences de contournement du gène majeur *pvr2³* et d'accumulation virale, avec mes données estimées de coefficient de sélection et de tailles efficaces de populations pour étudier l'influence de la dérive génétique, de la sélection et de l'accumulation virale sur la durabilité d'une résistance qualitative (article 3). Avant de présenter ces articles, je ferai dans le chapitre suivant une synthèse bibliographique sur la modélisation en épidémiologie, avec un accent particulier sur l'épidémiologie végétale.

Chapitre 2

Épidémiologie, modélisation et application aux agroécosystèmes

2.1 Définitions

L'épidémiologie est définie par Madden et al. (2007) comme l'étude du changement de l'intensité d'une maladie dans une population d'hôtes dans le temps et dans l'espace. Ainsi, l'épidémiologie s'intéresse à (i) des processus dynamiques, (ii) concernant les maladies (et pas uniquement les agents pathogènes ou leurs hôtes), (iii) dont l'ampleur peut être caractérisée par différents paramètres tels que le nombre de cas ou la durée d'une épidémie (Débarre and Gandon, 2009), (iv) dont l'étendue se situe au niveau d'une population d'hôtes, et (v) qui se déroule dans le temps et dans l'espace. Pour clarifier les choses sur le terme de maladie, un parasite qui colonise un hôte conduit à son infection. Si cette infection entraîne des symptômes néfastes pour l'hôte, alors on dit que l'hôte est malade. Sans symptômes néfastes identifiés, on ne peut pas parler de maladie. Enfin, un agent pathogène est un parasite qui induit une maladie chez un hôte (Begon et al., 2006).

Ainsi, l'*épidémiologie* se concentre sur la dynamique du nombre d'hôtes infectés dans une population (dans le temps et dans l'espace). Une deuxième composante est également importante dans la dynamique des maladies infectieuses. Il s'agit de l'*évolution* de l'agent pathogène. Débarre and Gandon (2009) ont défini l'évolution dans ce contexte comme les changements de la maladie, suite à l'apparition de nouvelles souches qui ont des propriétés différentes, telles que pour la virulence, l'agressivité ou la transmission, face aux résistances aux médicaments, aux vaccins, etc.. Pour ce qui concerne la phytopathologie, nous pouvons ajouter à cette liste l'évolution face à une résistance génétique chez la plante hôte. Les deux

processus d'évolution chez les agents pathogènes et d'épidémies chez les hôtes interagissent l'un avec l'autre, et il est nécessaire d'acquérir des connaissances sur les deux phénomènes pour avoir une compréhension globale d'une maladie infectieuse. La discipline qui étudie ces deux processus simultanément s'appelle *épidémiologie évolutive* (ou écologie évolutive des maladies infectieuses), elle s'intéresse à la dynamique à court et long terme de la maladie (Débarre and Gandon, 2009).

Dans la suite de ce chapitre, je parcourrai les grandes lignes de l'histoire des études épidémiologiques des agents pathogènes, puis des différentes caractéristiques qui distinguent les parasites, qui peuvent être déterminantes pour la définition du cadre d'étude épidémiologique. J'introduirai ensuite un exemple de modèle épidémiologique simple que je détaillerai pour comprendre le fonctionnement et l'intérêt de ce type de modèles. Enfin, j'aborderai différents types de modèles développés en épidémiologie végétale pour une gestion durable de la durabilité de la résistance des plantes aux agents pathogènes.

2.2 Épidémiologie et caractéristiques des parasites

2.2.1 Épidémiologie animale et humaine

Les premiers écrits de réflexions concernant les maladies infectieuses remontent à avant notre ère. Notamment, le livre de l'Exode (datation historique 8^{ème} - 3^{ème} siècle avant notre ère) décrit les 10 plaies amenées par Moïse en Égypte, parmi lesquelles figurent des épidémies touchant les hommes, le bétail et les cultures. Le terme plaie, *plague* en anglais, qui signifie également tourmenter (mais aussi peste), est d'ailleurs utilisé à cause d'une croyance commune que les épidémies représentent un châtement divin pour les pécheurs (Brauer and Castillo-Chávez, 2001). Hippocrate (460 - 370 avant notre ère), fondateur de la médecine grecque, a décrit avec précision certaines maladies chez les humains, telles que les oreillons et le paludisme, et suggérait l'importance de l'environnement sur l'émergence d'une maladie (Madden et al., 2007). Vers la même époque, Aristote (384 - 322 avant notre ère) a avancé l'idée que des créatures vivantes invisibles pouvaient être des agents de maladies. L'hypothèse d'Aristote fut prouvée au 17^{ème} siècle par le néerlandais Leuwenhoek (1632 - 1723) – considéré comme le père de la microbiologie – grâce aux premiers microscopes. En 1840, Jacob Henle jeta les premières pierres de la théorie microbienne, qui fut ensuite développée par Robert Koch, Joseph Lister et Louis Pasteur entre la fin du 19^{ème} siècle et le début du 20^{ème} siècle (Brauer and Castillo-Chávez, 2001).

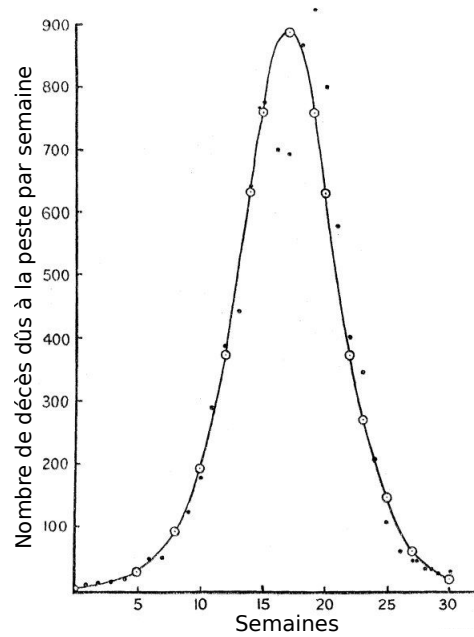


FIGURE 2.1 – **Ajustement du modèle de Kermack et McKendrick aux données de peste à Bombay entre le 17 décembre 1905 et le 21 juin 1906.** Nombre de personnes décédées par semaine à cause de la peste au cours du temps (en semaines). Les points noirs correspondent aux données observées, et les points blancs reliés par une courbe noire aux prédictions du modèle. Source : (Kermack and McKendrick, 1927).

Les premiers développements en modélisation mathématique des maladies transmissibles ont souvent été réalisés par des médecins de la santé publique (Brauer and Castillo-Chávez, 2001). La première étude mathématique connue pour l'épidémiologie a été menée par Daniel Bernoulli en 1766, défendant la pratique de la vaccination contre la variole (Bernoulli and Blower, 2004). Les premières contributions à la modélisation épidémiologique moderne ont été fournies par P.D. En'ko entre 1873 et 1894 (Dietz, 1988), et les fondations de l'approche compartimentale en épidémiologie ont été apportées par des médecins tels que R.A. Ross, W.H. Hamer, A.G. McKendrick et W.O. Kermack entre 1900 et 1935, ainsi que d'importantes contributions d'un point de vue statistique de J. Brownlee (Brauer and Castillo-Chávez, 2001). Ronald Ross est notamment connu pour ses travaux sur le paludisme, lui ayant valu le second prix Nobel de Médecine pour sa démonstration des dynamiques de transmission du paludisme entre moustiques et humains. Ronald Ross et William Hamer ont introduit les lois d'action de masse pour expliquer les dynamiques épidémiques. De leurs côtés, W.O. Kermack et A.G. McKendrick ont conçu, entre autres, la formulation d'un modèle simple capable de reproduire le comportement de nombreuses épidémies, telles que la peste (Grande peste de Londres, 1665 - 1666, tuant environ 20% de la population, et Bombay, 1905 - 1906, voir Fig. 2.1) ou le choléra (Londres, 1865). Ce comportement caractéristique consiste en une croissance exponentielle

suivie d'un déclin (Brauer et al., 2008).

2.2.2 Épidémiologie végétale

Les premiers écrits concernant les maladies des plantes sont datés plus tardivement que pour les maladies humaines. Ce sont Théophraste (372 - 287 avant notre ère) et plus tard Pline l'Ancien (23 - 79) qui en sont à l'origine, et ils ont également mis en avant le rôle de l'environnement dans les épidémies (Madden et al., 2007).

Zadoks and Koster (1976) attribuent la première documentation d'une étude épidémiologique chez les plantes à Duhamel de Monceau (1728), un botaniste et chimiste français qui a identifié sur les bulbes du crocus (*Crocus sativus*), dont la fleur produit le safran, une maladie causée par un champignon (*Helicobasidium purpureum*) transmis par le sol, et a pu fournir des méthodes de contrôle basées sur sa connaissance des épidémies. Il a même prouvé que la maladie était causée par une entité biologique (Madden et al., 2007). Dans le probablement premier manuel de pathologie végétale, Kühn (1858) relata l'importance des épidémies dans les cultures, et les compara aux épidémies chez les humains et chez les autres animaux.

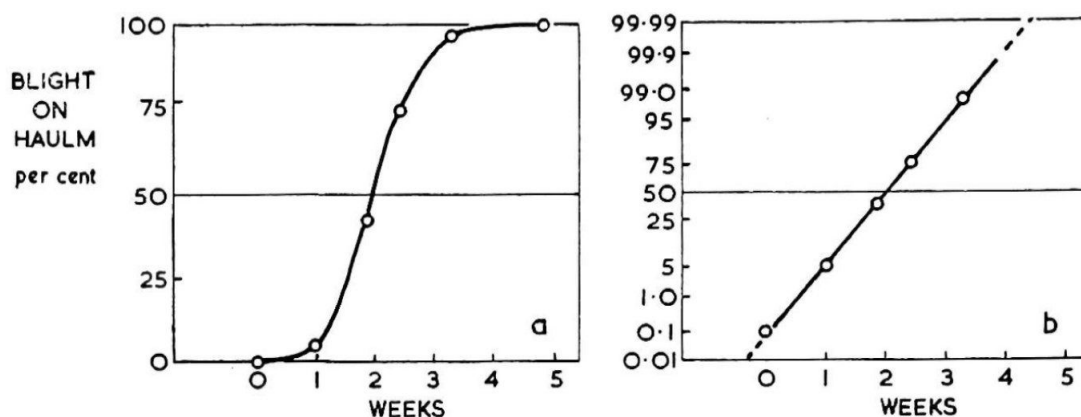


FIGURE 2.2 – Courbe de progression typique du mildiou chez la pomme de terre (notation basée sur le pourcentage de surface foliaire infectée) au cours des semaines, (a) dans une échelle normale, (b) dans une échelle Probit. Source : (Large, 1952).

Une transition importante pour l'épidémiologie végétale s'opéra vers le milieu du 20^{ème} siècle, cette discipline initialement qualitative s'ouvrant à l'aspect quantitatif. Large (1952) a fait partie des pionniers de cette transition, en montrant l'intérêt des courbes de progression de maladies (Fig. 2.2), notamment pour prédire les pertes de rendement (Colhoun, 1993). Il a montré que les courbes de progression du pourcentage de surface foliaire de pomme de terre détruit par le mildiou au cours du temps étaient sigmoïdales, et pouvaient être réduites à des droites grâce à une transformation Probit. Il a également montré que les épidémies dues

au mildiou pouvaient être complètement décrites sachant le temps de demie-vie¹ ainsi que la pente de la droite de décroissance de surface foliaire saine. Un peu plus tard, un cultivateur de pommes de terre en Afrique du Sud, van der Plank (1960), écrivit un chapitre remarquable, intitulé "Analyse d'épidémies", dans lequel il fit part de l'idée radicale (du moins pour les phytopathologistes de l'époque) que les épidémies pouvaient être modélisées, et qu'une analyse basée sur des modèles était essentielle pour comprendre les processus de progression d'une maladie et pour établir des stratégies de contrôle (Madden et al., 2007). L'écriture de ce chapitre fut suivie de la rédaction d'autres livres sur l'épidémiologie, la résistance des hôtes et les interactions plante - agents pathogènes, parmi lesquels son premier livre (van der Plank, 1963) est d'une importance majeure. Il y développa une théorie de l'épidémiologie végétale basée sur des processus de dynamique des populations, et sur des données empiriques issues de la littérature (Madden et al., 2007).

Il existe un lien fort entre les modèles épidémiologiques pour la médecine et pour les plantes (Madden et al., 2007; Ohtsuki and Sasaki, 2006). En épidémiologie humaine ou animale, la maladie se répand à cause de la mobilité des individus et de leur probabilité de rencontre. En épidémiologie végétale, les plantes sont fixes mais la maladie se répand à cause de la motilité des inocula² produits par les agents pathogènes (transport par le vent, par l'eau, par un vecteur, etc.) (Madden et al., 2007). Le paramètre commun à ces deux disciplines est au final le taux de transmission du parasite.

De manière plus générale, d'après McNeill (1976), c'est l'émergence des villes et le développement de l'agriculture, conduisant à une densification des populations humaines, qui ont permis à des micro-parasites tels que les virus et les bactéries de se maintenir sans hôtes intermédiaires. Ainsi, les maladies directement transmissibles, telles que la rougeole, la grippe ou la tuberculose, sont le résultat de notre vie moderne (Brauer and Castillo-Chávez, 2001). D'autres maladies, transmises indirectement par des vecteurs, tel que le paludisme, sont également d'une importance majeure chez les humains.

2.2.3 Micro- et macro-parasites

Les micro-parasites les plus connus sont les bactéries et les virus. Chez les animaux, on trouve par exemple le virus de la rougeole (*Measles virus*) ou la bactérie de la typhoïde (*Salmonella enterica*) (Begon et al., 2006). Chez les plantes, quelques exemples concernent

1. Temps au bout duquel il ne reste plus que 50% de surface foliaire saine.

2. Propagules du parasite capables de contaminer une plante hôte.

le virus de la mosaïque du concombre (CMV), ou le virus de la maladie bronzée de la tomate (TSWV), infectant plus de 1000 espèces de plantes, ou encore la bactériose du riz (*Xanthomonas oryzae* pv *oryzae*) pouvant induire jusqu'à 70% de pertes de rendement (Begon et al., 2006; García-Arenal and McDonald, 2003; McDonald and Linde, 2002a). Le troisième groupe de micro-parasites infectant les animaux est constitué des protozoaires, tels que les trypanosomes, à l'origine de la maladie du sommeil, et les espèces de *Plasmodium* causant le paludisme) (Begon et al., 2006). Chez les plantes hôtes, certains champignons microscopiques (unicellulaires) sont assimilables à des micro-parasites. Les micro-parasites se caractérisent chez les animaux et les plantes par de petites tailles, des temps de génération courts, des taux de reproduction dans l'hôte très élevés, et une tendance à induire une immunité contre les réinfections chez les hôtes survivant au premier contact (Anderson and May, 1979). La durée de l'infection est typiquement courte comparée à l'espérance de vie de l'hôte, et est donc généralement de nature transitoire (sauf si l'infection cause le décès de l'hôte) (Begon et al., 2006).

Un groupe de macro-parasites majeur des animaux est constitué des helminthes, qui sont des vers parasites (Begon et al., 2006). Par exemple, les nématodes des intestins des humains, tous transmis directement, sont certainement les parasites des intestins humains les plus importants, en termes de nombre de personnes infectées comme pour leur potentiel à causer des problèmes de santé. Des macro-parasites avec cycle de vie indirect³ sont également importants d'un point de vue médical. En font partie les ténias (vers solitaires), parasites des intestins durant leur stade adulte, absorbant les nutriments directement à travers les parois de leur corps et produisant des œufs relâchés dans les matières fécales. Durant leur stade larvaire, les ténias passent par un ou deux hôtes intermédiaires avant d'infecter leur hôte définitif. De plus, certains arthropodes (tels que le pou, la puce, la tique, la mite) ainsi que certains champignons attaquant les animaux font partie des macro-parasites. Les macro-parasites des plantes incluent les champignons macroscopiques (pluricellulaires), à l'origine du mildiou (moisissure), des rouilles et charbons, ainsi que les insectes formant des galles ou des galeries. Les macro-parasites montrent généralement des temps de génération beaucoup plus longs que les micro-parasites, et la multiplication directe dans l'hôte est soit absente, soit se passe à un taux réduit (Anderson and May, 1979). Les réponses immunitaires qu'ils induisent dépendent généralement du nombre de parasites dans un hôte donné, et ont tendance à être de courte durée. Leurs infections sont donc généralement de nature persistante, avec des hôtes continuellement réinfectés (Begon et al., 2006).

3. Parasites nécessitant au moins deux hôtes, d'espèces différentes, pour compléter leur cycle de vie.

2.2.4 Modes de transmission

Au-delà des distinctions entre micro- et macro-parasites, les parasites se distinguent par leur mode de transmission : directe ou indirecte (Anderson and May, 1979; Begon et al., 2006). La transmission directe peut se faire par contact, par inhalation ou par ingestion. Un cas spécial de transmission directe est celle du parent à l'enfant, appelée transmission *verticale*. Chez les plantes, cette transmission se fait par la graine (Astier et al., 2001). Par opposition, tous les autres cas de transmission sont appelés transmission *horizontale*. La transmission indirecte peut faire appel à un vecteur ou à une ou plusieurs espèces hôtes intermédiaires (cycle de vie indirect). Le terme vecteur fait référence à un organisme vivant transportant un parasite d'un hôte à un autre. Certains vecteurs n'ont que ce rôle de transporteur, d'autres sont également des hôtes intermédiaires dans lesquels le parasite se multiplie. Chez les plantes et leurs virus, la plupart des insectes vecteurs appartiennent à deux ordres d'insectes suceurs qui s'alimentent sur les plantes, les thysanoptères (thrips) et les hémiptères, dont font partie par exemple les pucerons (Ferreles and Raccach, 2015; Nault, 1997). On distingue ensuite différentes caractéristiques de modes de transmission chez ces insectes (Ferreles and Raccach, 2015). Je vais prendre l'exemple des pucerons, pour lesquels les différentes caractéristiques de transmission peuvent être observées. Selon le mode de transmission dit *non circulant*, le virus reste dans la double paire de stylets (pièce buccale) du puceron avant d'être introduit dans une nouvelle plante par piqûre. Ils restent généralement moins de 12 heures dans les stylets, et les temps d'acquisition et d'inoculation sont très courts (moins de 5 minutes) (Moury et al., 2007). On distingue deux types de transmissions non circulantes : *non persistante* et *semi persistante*, qui se différencient notamment par des durées d'acquisition, de transmission et de rétention un peu plus longues et une réussite d'inoculation un peu plus élevée pour le mode semi persistant (Ferreles and Raccach, 2015). Selon le mode de transmission dit *circulant*, ou *persistent*, les temps d'acquisition et d'inoculation sont plus longs, et le virus peut rester de 12 heures à la durée de vie du puceron à l'intérieur de celui-ci. Une période de latence entre l'acquisition et l'inoculation est généralement nécessaire, pouvant durer jusqu'à plusieurs jours (Moury et al., 2007).

2.2.5 Définition du cadre d'étude

Il est important d'avoir certaines informations basiques sur les caractéristiques du parasite étudié, telles que celles évoquées précédemment, pour développer un modèle épidémiologique cohérent avec le modèle biologique sous-jacent. Les modèles épidémiologiques s'intéressent aux dynamiques de transmission de parasites d'hôte en hôte (Begon et al., 2006; Brauer and

Castillo-Chávez, 2001). Le choix de l'unité épidémiologique va dépendre du parasite, mais aussi de la question à laquelle veut répondre l'investigateur (Brauer and Castillo-Chávez, 2001). En général, pour les micro-parasites, l'unité épidémiologique correspond au nombre d'hôtes infectés, les parasites étant en trop grand nombre au sein de l'hôte pour être dénombrés. Pour les macro-parasites, l'unité épidémiologique peut être le nombre de parasites ou le nombre d'hôtes infectés (Begon et al., 2006). La distinction entre micro- et macro-parasites n'est pas la seule à permettre de définir l'unité épidémiologique, les caractéristiques d'une maladie peuvent aussi servir de critère. Par exemple, dans le cas du paludisme, l'unité épidémiologique peut correspondre aux hôtes (humains et autres mammifères) et aux vecteurs (moustiques femelles). Dans le cas de la maladie de Chagas la maison est généralement adoptée comme unité épidémiologique (Brauer and Castillo-Chávez, 2001). Dernier exemple, pour le mildiou elle correspond généralement à la feuille ou à la surface de feuille donnant lieu à un cycle de réplication du mildiou à partir d'une seule spore. De manière plus générale, le choix de l'unité épidémiologique va déterminer le niveau de détails du modèle. Il y a un compromis entre les modèles simples, qui s'affranchissent de la plupart des détails et mettent l'accent sur le comportement qualitatif du modèle, et les modèles complexes, généralement conçus pour des situations spécifiques et mettant l'accent sur le comportement quantitatif à court terme (Brauer and Castillo-Chávez, 2001). Un modèle mathématique propose une représentation de la réalité, mais il ne pourra jamais en être une copie conforme à cause de sa trop grande complexité. La réalité étant complexe, pour concevoir un modèle il faut donc faire des choix et formuler un certain nombre d'hypothèses simplificatrices qui doivent être justifiées biologiquement. L'intérêt des modèles épidémiologiques, et plus généralement de l'application de modèles mathématiques pour la biologie, est de comprendre les mécanismes sous-jacents impliqués dans des processus biologiques, mais aussi de prédire l'évolution d'un système (Murray, 2002).

2.3 Introduction aux modèles à compartiments

2.3.1 Le modèle déterministe SIR de Kermack et McKendrick

Dans les modèles compartimentaux, caractéristiques de la vaste majorité des modèles épidémiologiques, la population étudiée est divisée en compartiments, avec des hypothèses sur la nature et le taux de transfert d'un compartiment à un autre (Brauer and Castillo-Chávez, 2001). Les maladies conférant une immunité ont une structure compartimentale différente de celles n'en conférant pas. Les maladies transmises par un vecteur constituant un hôte intermédiaire ont

encore une autre structure.

Je vais ici présenter une version simple et classique du modèle de Kermack and McKendrick (1927), appelé modèle SIR, constituant un bon exemple de modèle épidémiologique en équations différentielles ordinaires (*Ordinary Differential Equations*) (ODE) (Brauer and Castillo-Chávez, 2001; Brauer et al., 2008). Par souci de simplicité, j'appellerai ce modèle SIR le modèle de Kermack-McKendrick et je ne parlerai pas du modèle de base qui est plus compliqué. Ce modèle représente l'évolution au cours du temps du nombre de personnes infectées par une maladie directement transmissible dans une population fermée, c'est-à-dire sans migration, et à taille de population constante. Il repose sur trois hypothèses principales : (i) l'échelle de temps de la maladie est plus rapide que celle des naissances et morts naturelles, donc les effets démographiques non liés à la maladie sont négligés, (ii) il n'y a pas de période de latence entre l'acquisition et la transmissibilité de l'agent pathogène, et (iii) la période de contagion est la même que celle de l'infection. Le modèle se compose de trois compartiments :

- *Susceptible* (S) : les individus sains, qui n'ont pas d'immunité contre le parasite et qui peuvent donc devenir infectés suite au contact avec ce dernier,
- *Infected* (I) : les individus infectés, qui sont également infectieux et peuvent donc transmettre le parasite aux individus sains par contact,
- *Removed* (R) : les individus qui ne sont plus infectieux, soit par isolation du reste de la population, par immunisation contre l'infection, par guérison de la maladie avec une immunité acquise contre la réinfection, ou encore par mort causée par la maladie.

Les nombres d'individus dans chacun de ces compartiments correspondent aux variables d'état du système et sont dénotés S , I et R . La taille totale de la population s'écrit $N = S + I + R$. Le modèle est représenté de manière schématique dans la figure 2.3. On peut y observer que la maladie conduit inévitablement à l'infection puis à la non-contagiosité des personnes saines (sens des flèches qui vont vers le compartiment I puis R). Les personnes saines deviennent infectées à un taux β (per capita) lorsqu'elles sont en contact avec des personnes infectées. Les personnes infectées deviennent non-infectieuses à un taux γ . La période moyenne d'infectiosité vaut $1/\gamma$.

Le système d'ODE non linéaires correspondant s'écrit comme suit :

$$\dot{S} = -\beta SI \tag{2.1}$$

$$\dot{I} = \beta SI - \gamma I \tag{2.2}$$

$$\dot{R} = \gamma I, \tag{2.3}$$

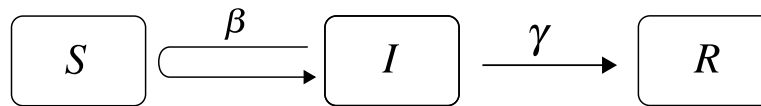


FIGURE 2.3 – **Représentation schématique du modèle SIR de Kermack and McKendrick (1927)**. Ce modèle est à trois compartiments : individus sains S , individus infectés I et individus qui ne sont plus infectieux R . Les taux de transfert entre compartiments sont : β , le taux d'infection des personnes saines en contact avec des personnes infectées, et γ , le taux de passage à l'état non-infectieux des personnes infectées.

définissant la variation des trois variables d'état au cours du temps. Dans ce modèle, R est déterminé dès que S et I sont connus, tel que $R = N - S - I$. L'équation 2.3 définissant la variation de R est donc facultative, les équations 2.1 et 2.2 suffisent à décrire le système. Les taux de transfert entre compartiments β et γ sont exprimés par unité de temps et par individu. Le terme βSI , bilinéaire, vient de la loi d'action de masse. Le paramètre β , coefficient de transmission, est en fait en partie un produit entre (i) la probabilité p qu'un contact pouvant conduire à la transmission du parasite le transmette effectivement et (ii) le taux de contact c entre un individu sain S et tous les autres hôtes (ou vecteurs). La proportion d'individus infectés étant I/N , le taux de production de nouvelles infections vaudrait donc $cpSI/N$ (Begon et al., 2006). On peut en déduire le paramètre $\beta = cp/N$. La définition de β suppose que les tailles des compartiments sont suffisamment grandes pour que le mélange des individus soit homogène (Brauer et al., 2008), et donc le taux de contact entre un individu sain et un individu infecté, cI/N , est constant dans toute la population.

Il n'y a pas de solution analytique à ce système d'ODE (à cause du terme bilinéaire), mais nous savons qu'un équilibre sans maladie satisfait $(S, I, R) = (N, 0, 0)$. Si un individu devient infecté, il est attendu qu'il infecte d'autres individus à un taux βN pendant sa période infectieuse prévisionnelle $1/\gamma$. Ainsi, le premier individu infecté transmettra en moyenne la maladie à $R_0 = \beta N/\gamma$ autres individus. Le paramètre R_0 est le taux de reproduction de base, qui est un paramètre très important en épidémiologie. Il définit le seuil épidémiologique de la maladie. Si $R_0 < 1$, la maladie s'éteint car les individus infectés n'arrivent pas à infecter suffisamment d'autres personnes. Si $R_0 > 1$, il y a une épidémie car chaque individu infecté arrive à infecter en moyenne au moins une personne (Brauer and Castillo-Chávez, 2001; Brauer et al., 2008).

Pour certaines maladies, une période de latence est observée entre l'infection d'un individu (acquisition du parasite) et sa contagiosité (transmissibilité du parasite) (Brauer et al., 2008). Pour modéliser ce phénomène, il convient d'ajouter un compartiment E (*Exposed*, exposés) entre les compartiments S et I dans le schéma de la figure 2.1. Cela ajoute donc une équation

au système, définissant la variation de la variable d'état E au cours du temps. D'autres configurations peuvent être ainsi ajoutées au modèle basique présenté, telles qu'un traitement, par exemple par vaccination, une distinction des individus par classe d'âge, ayant un impact sur la capacité du parasite à les infecter et sur leur contagiosité, etc..

2.3.2 Version stochastique d'un modèle SIS simple

Le modèle de Kermack et McKendrick, en ODE, suppose que le processus épidémique est *déterministe*, c'est-à-dire que le comportement de l'épidémie est complètement déterminé par son histoire et par les règles régissant le modèle (Brauer et al., 2008). D'autres types de modèles, dit *stochastiques*, utilisent des concepts probabilistes et définissent une distribution de comportements possibles. Ces modèles sont particulièrement utiles en épidémiologie. Ils ont par exemple été utilisés pour modéliser l'effet de la dérive génétique sur le polymorphisme chez les agents pathogènes et leurs plantes hôtes évoluant selon le modèle gène-pour-gène (Salathé et al., 2005), pour étudier le rôle de l'évolution des agents pathogènes sur l'émergence de maladies infectieuses en provenance d'un réservoir (animaux) chez les hommes (Antia et al., 2003), ou encore pour étudier l'effet de la stochasticité démographique (Encadré 1.5) et environnementale sur l'évolution de populations d'agents pathogènes et sur la durabilité de la résistance des plantes (Lo Iacono et al., 2013). Il existe, en plus des modèles individus-centrés (*Individual Based Model*) (IBM), trois formulations de modèles stochastiques : les chaînes de Markov en temps discret (*Discrete Time Markov Chain*) (DTMC), les chaînes de Markov en temps continu (*Continuous Time Markov Chain*) (CTMC) et les équations différentielles stochastiques (*Stochastic Differential Equations*) (SDE). Ces trois formulations diffèrent par les hypothèses sous-jacentes à la modélisation du temps et des variables d'état. Dans un modèle DTMC, le temps et les variables d'état sont discrets, dans un modèle CTMC le temps est continu et les variables d'état sont discrètes, et dans un modèle SDE les deux sont continus. Enfin, par définition, un modèle IBM s'intéresse aux trajectoires de chaque individu séparément, par exemple à l'évolution de l'état d'une plante hôte d'un agent pathogène au cours du temps. La différence la plus importante entre des modèles épidémiologiques déterministes et stochastiques réside dans leurs dynamiques asymptotiques. Des simulations d'un modèle stochastique pourront converger vers un état sans maladie même si la simulation déterministe correspondante converge vers un équilibre endémique (Brauer et al., 2008). Dans la suite, je ne présenterai que les CTMC, qui seront utilisées dans le modèle épidémiologique développé dans le cadre de cette thèse.

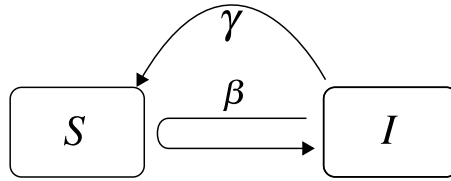


FIGURE 2.4 – Représentation schématique du modèle SIS issu de Kermack and McKendrick (1927). Ce modèle est à deux compartiments : individus sains S et individus infectés I . Les taux de transfert entre compartiments sont : β , le taux d'infection des personnes saines en contact avec des personnes infectées et γ , le taux de rétablissement des personnes infectées.

Modèle SIS déterministe

Par souci de simplicité, je vais prendre l'exemple d'un modèle SIS, c'est-à-dire Sain-Infecté-Sain, à deux compartiments. Cela se fait très simplement à partir du modèle déterministe précédent (Eq. 2.1-2.3), en faisant disparaître le compartiment R . Sous l'hypothèse de taille de population constante ($N = S + I$), un individu sortant du compartiment I retourne dans le compartiment S . Ainsi, cela suppose que les individus infectés peuvent guérir, mais ne seront pas immunisés contre un nouveau contact avec l'agent pathogène. Une représentation schématique de ce modèle est donnée dans la figure 2.4. Le paramètre β a le même sens que précédemment alors que le paramètre γ correspond maintenant aux taux de rétablissement des individus infectés.

Le nouveau système d'ODE non linéaires correspondant au modèle SIS s'écrit comme suit :

$$\dot{S} = -\beta SI + \gamma I \quad (2.4)$$

$$\dot{I} = \beta SI - \gamma I. \quad (2.5)$$

Modèle SIS stochastique

Dans un modèle SIS stochastique de type CTMC, le temps est continu, et les variables d'état sont des variables aléatoires discrètes, telles que $S(t), I(t) \in \{0, 1, 2, \dots, N\}$. Puisque la taille de population N est constante, je vais me focaliser sur la variable aléatoire discrète liée au nombre d'individus infectés au temps t , $\{I(t)\}, t \in [0, +\infty[$. La fonction de probabilité associée est $p(t) = (p_0(t), \dots, p_N(t))^T$, où $p_i(t)$ correspond à la probabilité d'observer i individus infectés au temps t . Puisque $p(t)$ correspond à une distribution de probabilité, on doit avoir :

$$\sum_{i=1}^N p_i(t) = 1, \forall t \in [0, +\infty[.$$

Le vecteur $p(t)$ peut être obtenu par la résolution d'un système d'équations différentielles ordinaires (Brauer et al., 2008).

Une des manières de construire un modèle CTMC est de définir des taux de transition d'un état vers un autre, par unité de temps. Sur la base de l'ODE (2.5), les taux de transitions du nombre d'individus infectés $\pi_{i,j}$ de l'état i vers l'état j sont :

$$\begin{aligned}\pi_{i,i+1} &= \beta i(N - i) \\ \pi_{i,i-1} &= \gamma i\end{aligned}\tag{2.6}$$

Un processus stochastique possède la *propriété de Markov*, c'est-à-dire que l'état à venir de la population ne dépend que de son état actuel, pas des états passés. Autrement dit, la probabilité de transition au temps t_{n+1} ne dépend que du temps présent t_n .

Les transitions décrites peuvent être assimilées à des processus de naissance et de mort, du point de vue des individus infectés. Ainsi, le premier évènement, pour lequel $j = i + 1$, correspond à la naissance d'un individu infecté, et le second, pour lequel $j = i - 1$, à sa mort.

Modélisation du temps

Par construction classique d'un modèle CTMC, le temps d'attente entre deux évènements suit une loi exponentielle. Les évènements sont issus d'un processus de poisson, ils se produisent de manière continue et indépendante, à un taux moyen constant égal à la somme des taux de transition individuels. Si, au temps présent t , i individus sont infectés, alors le taux de naissance correspondant vaut $\beta i(N - i)$ et le taux de mort vaut γi , tel que décrit dans le système 2.6. Le temps d'attente s'écrit alors :

$$\Delta t \sim \text{Exp}(\beta i(N - i) + \gamma i),$$

le paramètre de la loi exponentielle correspondant à la somme des taux de naissance et de mort.

Une fois que le temps de cet évènement est atteint, il faut déterminer lequel a lieu. Cela se fait à l'aide de probabilités construites à partir des taux de transition :

$$P_{i,i+1} = \frac{\pi_{i,i+1}}{\pi_{i,i+1} + \pi_{i,i-1}} \quad \text{et} \quad P_{i,i-1} = \frac{\pi_{i,i-1}}{\pi_{i,i+1} + \pi_{i,i-1}}$$

La simulation d'une instance d'un modèle CTMC se fait alors via l'algorithme de Gillespie (Wilkinson, 2006), comme suit.

1. Initialisation du système : donner une valeur à $I(0) = i$ (par définition, $S(t) = N -$

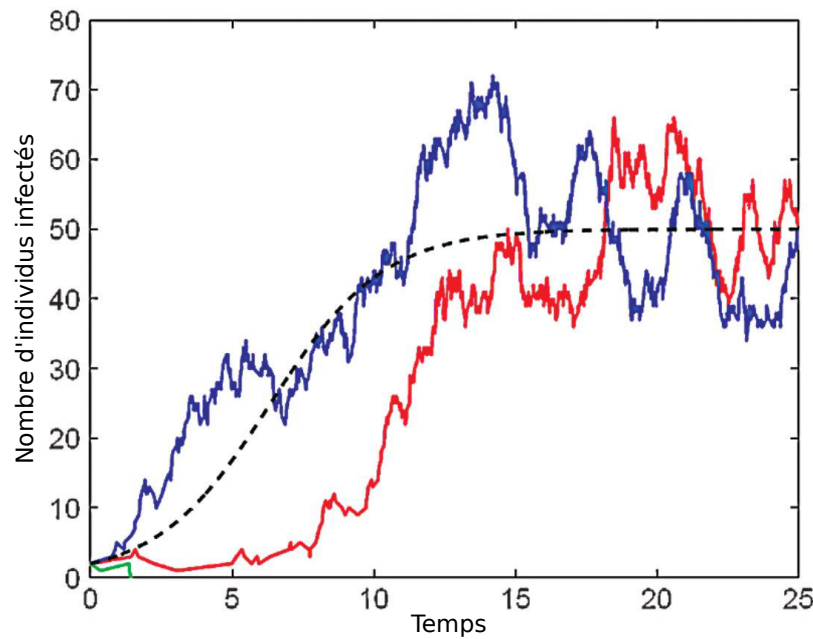


FIGURE 2.5 – Représentation de trois trajectoires évolutives du nombre d’individus infectés obtenues avec le modèle épidémique SIS de type CTMC (en bleu, rouge et vert), ainsi que la solution déterministe (courbe en pointillés noirs). Les paramètres ont été fixés à : $\gamma = 0.5$, $\beta = 1$, $N = 100$ et $I(0) = 2$. Source : (Brauer et al., 2008).

$I(t)$,

2. Calcul du temps d’attente Δt au prochain évènement de naissance ou de mort,
3. Déterminer le type du prochain évènement (par un processus de Bernoulli),
4. Mettre à jour le nombre d’individus infectés, i.e. $I(t + \Delta t) = i + k$, avec $k = -1$ ou 1 .
5. Recommencer depuis l’étape 2.

Ainsi, pour chaque simulation de cet algorithme, on obtient une courbe d’évolution du nombre d’individus infectés au cours du temps. Pour avoir une idée du comportement moyen du système, il faut généralement le simuler un grand nombre de fois. À titre d’illustration, trois trajectoires évolutives de ce modèle SIS ont été représentées dans la figure 2.5, avec la courbe d’évolution du modèle déterministe correspondant. On voit que deux des trajectoires suivent à peu près la courbe déterministe, alors que la troisième (en vert) arrive rapidement à un équilibre sans maladie. Comme je l’ai évoqué précédemment, ce cas représente une des différences les plus importantes entre les modèles épidémiologiques déterministes et stochastiques : les dynamiques asymptotiques peuvent être différentes entre les deux types de modèles (Brauer et al., 2008).

Avec des modèles simples, nous avons pu voir comment sont construits les modèles déterministes et stochastiques, et la logique sous-jacente. Je vais maintenant évoquer diverses applications de modèles épidémiologiques pour la gestion de la durabilité des plantes, et voir

les spécificités de ces modèles.

2.4 Applications pour la gestion de la durabilité de la résistance des plantes

Depuis les travaux de van der Plank, un nombre considérable d'études faisant appel à la modélisation ont été réalisées pour étudier les épidémies d'agents pathogènes dans les cultures. Notamment, van der Plank (1963) avait clairement identifié l'importance de la relation entre maladie et rendement agricole, en analysant les conséquences des épidémies (Madden et al., 2007). Les années 70 et 80 furent très productives pour les recherches sur ce sujet, avec entre autres des contributions de W. Clive James et Paul Teng (e.g. James (1974); Teng and Gaunt (1980)). Les informations tirées de ces recherches sont primordiales face au poids économique des pertes de rendement dues aux maladies dans les cultures. Elles permettent notamment d'identifier l'effet de stratégies de contrôle sur les rendements.

La résistance des plantes aux agents pathogènes (voir partie 1.2) fait partie des stratégies de contrôle. Cette stratégie est étudiée depuis de nombreuses années car la durabilité des résistances est souvent faible (Kiyosawa, 1982; Leonard, 1977; van der Plank, 1963, 1968). Dans la suite, nous allons voir différentes approches de modélisation développées pour étudier les épidémies d'agents pathogènes dans les cultures, la plupart faisant intervenir des variétés résistantes. De telles études peuvent focaliser leur critère de durabilité sur les rendements agricoles, mais aussi sur la fréquence des variants adaptés à la résistance, indicatrice de l'efficacité de la résistance (van den Bosch and Gilligan, 2003).

2.4.1 Caractéristiques liées à l'épidémiologie dans les agro-écosystèmes

Je vais commencer par donner quelques exemples d'études en lien avec des caractéristiques qui peuvent être retrouvées dans les agro-écosystèmes, à savoir la saisonnalité des cultures, leur dimension spatiale, et différents effets stochastiques affectant les épidémies.

Saisonnalité

La saisonnalité, qui est une composante de la plupart des cultures dans les climats tempérés, a été formalisée pour la modélisation grâce aux modèles semi-discrets (Fig. 2.6) (Mailleret and Lemesle, 2009). Ce sont des systèmes dynamiques hybrides qui suivent des

dynamiques continues la plupart du temps, mais passent régulièrement par des changements discrets. Ils sont utilisés dans différents domaines de recherche, tels que l'épidémiologie humaine (campagnes de vaccination, Agur et al. (1993)) ou végétale (saisonnalité dans les cultures, Shaw (1994)), en médecine (traitement médicamenteux, Panetta (1996, 1998)), ou encore en dynamique des populations (programmes de protection biologique introduisant de manière répétée une espèce auxiliaire pour le contrôle d'une espèce nuisible, Mailleret and Grogard (2006, 2009)).

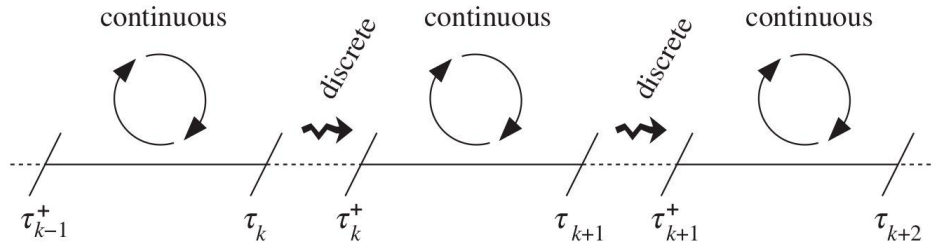


FIGURE 2.6 – **Illustration graphique d'un modèle semi-discret.** L'axe horizontal représente le temps. Lorsque le temps t est entre deux instants τ_{k-1}^+ et τ_k , le système suit des dynamiques continues ; il subit des changements discrets aux instants $t = \tau_k$. Source : Mailleret and Lemesle (2009).

Dans un paysage agricole, la saisonnalité a pour conséquence la disparition des cultures pendant une partie de l'année, entre la récolte et le semis de la saison culturale suivante. En conséquence, les agents pathogènes subissent des goulets d'étranglement de manière répétée, à chaque fin de saison de culture marquée par la disparition de leur population d'hôtes (Mailleret and Lemesle, 2009). Pendant les inter-saisons, les agents pathogènes prennent généralement des formes spéciales pour la survie, résistantes au froid (Mailleret et al., 2012). Les agents pathogènes peuvent être alors vus comme des inocula primaires car ils vont générer les infections primaires à la saison de culture à venir, suivies par les infections secondaires d'hôte à hôte. Ces infections secondaires seront à leur tour sources d'inocula primaires hébergés par les plantes sauvages avoisinantes.

Le premier modèle proposé en épidémiologie pour les maladies des plantes tenant compte des goulets d'étranglement saisonniers est attribué à Shaw (1994), inspiré du modèle proposé par Levins and Wilson (1980). Shaw (1994) a montré des dynamiques très complexes dues à ces processus saisonniers (Mailleret and Lemesle, 2009). Des études qui ont suivi se sont notamment intéressées à la persistance ou à l'extinction des populations d'agents pathogènes dans de tels systèmes (e.g. Gubbins and Gilligan, 1997a,b; Madden and van den Bosch, 2002; Truscott et al., 1997), ou à la différence entre un système semi-discret ou continu sur l'issue de la compétition entre deux virus pour une espèce de plante (Zhang and Holt, 2001). Zhang and Holt (2001)

ont montré qu'un principe d'exclusion compétitive s'appliquait dans le modèle continu alors que ce n'était pas toujours le cas dans le modèle semi-discret. Plus récemment, Mailleret et al. (2012) ont étudié l'existence ou non d'exclusion compétitive à la suite des infections primaires en début de saison de culture, mais cette fois-ci selon que les taux d'infection primaires étaient densité-dépendants ou non. Ils ont montré que le principe d'exclusion compétitive n'était pas toujours conservé lorsqu'on passait d'un type de modèle à l'autre.

De manière plus générale, un certain nombre d'études se sont attachées à mieux comprendre les conditions d'hivernage des agents pathogènes et les conséquences de cet hivernage sur les épidémies à la saison de culture suivante, mais il reste encore des choses à apprendre (Burdon and Thrall, 2008). Parmi les études expérimentales sur le sujet figure celle de Montarry et al. (2007), qui se sont demandés s'il existait un compromis évolutif entre l'agressivité de souches de *Phytophthora infestans* sur pomme de terre et leur capacité de survie à l'hiver. Les données collectées n'allaient pas dans ce sens, mais ils ont montré que les tubercules de pomme de terre infectées germaient plus tôt à la saison de culture suivante que celles non infectées. Suffert et al. (2011) ont mis en évidence que le mécanisme le plus important chez *Mycosphaerella graminicola* pour les stades précoces des épidémies et la transmission de la maladie d'une année à la suivante était l'infection par les ascospores⁴, dispersées par le vent depuis des débris de blé infectés distants ou proches. Papaïx et al. (2015) ont mené une étude théorique sur l'effet de la distribution spatiale et temporelle des patchs de végétation sauvage résiduelle implantés dans un paysage agricole sur l'émergence et la spécialisation d'un agent pathogène à une variété cultivée. Ils ont montré que les structures du paysage favorisant des grandes tailles de populations d'agents pathogènes dans les plantes sauvages facilitaient l'émergence de l'agent pathogène dans les cultures. En contre-partie, de telles structures du paysage réduisent le potentiel évolutif de l'agent pathogène vers une adaptation à la variété cultivée. Dans leur revue, Burdon and Thrall (2008) évoquent différents points clés sur l'hivernage, les plantes sauvages et leurs conséquences sur les populations d'agents pathogènes. Notamment, les plantes sauvages peuvent constituer un 'pont vert' entre la récolte d'une culture et le semis de la saison suivante, soit en tant qu'hôtes alternatifs des agents pathogènes (multiplication dans l'hôte), soit comme réservoir d'inoculum. En Australie, les différentes variétés d'avoine cultivées et sauvages interagissent avec leurs agents pathogènes de la rouille, par flux de gènes. Les études de ces interactions montrent que les hôtes sauvages contribuent à (i) sélectionner des nouvelles mutations, ainsi qu'à (ii) maintenir de la diversité génétique dans les populations d'agents pathogènes des plantes cultivées. Toutes ces études permettent une modélisation plus précise

4. Spore contenue dans un asque (cellule reproductrice).

des périodes d'inter-saisons et de leur conséquence sur les populations d'agents pathogènes. Elles sont donc d'une importance majeure pour la conception de modèles semi-discrets focalisés sur les épidémies dans les paysages agricoles sur plusieurs saisons de cultures.

Modèles spatialisés

Les modèles classiques tels que les modèles SIS ou SIR décrits précédemment sont implicitement spatialisés, sous l'hypothèse que tous les individus sont mélangés de manière homogène et qu'ils sont tous équidistants puisqu'ils ont tous le même taux de contact. Différents modèles se sont également attachés à intégrer explicitement la dimension spatiale dans leurs modèles épidémiologiques, pour étudier son effet sur l'émergence, l'extinction ou la persistance des ravageurs des cultures (e.g. Donalson and Nibset, 1999; Fahrig, 1997; Papaïx, 2011; Papaïx et al., 2013, 2014, 2015; Vacher et al., 2003).

Le modèle de Papaïx et al. (2015), évoqué précédemment dans la partie Saisonnalité, constitue un exemple de modèle spatialisé. Je prendrai ici comme exemple un autre article du même groupe d'auteurs. Dans cet article, Papaïx et al. (2014) ont étudié l'effet du niveau d'aggrégation des hôtes à l'échelle du paysage agricole sur le contrôle d'une maladie foliaire due à un agent pathogène transmis par voie aérienne. Le paysage était composé de deux variétés, une sensible et une résistante. Différents niveaux de résistance de cette dernière variété ont été testés, définissant des variétés porteuses de résistantes soit quantitatives soit qualitatives. Papaïx et al. (2014) ont montré que des paysages où les deux variétés étaient mélangées finement (faible niveau d'aggrégation) était plus efficace pour le contrôle de la maladie lors du déploiement d'une variété résistante qualitative. À l'inverse, avec le déploiement d'une variété résistante quantitative, leur modèle a montré qu'il pouvait être plus efficace d'aggréger les hôtes de la même variété dans différentes régions, dépendamment de la proportion de plantes résistantes et de leur niveau de résistance.

Stochasticités

Des modèles stochastiques ont aussi été utilisés, par exemple pour modéliser les effets aléatoires de l'environnement sur les épidémies ou les fluctuations aléatoires dans les dynamiques démographiques des populations d'agents pathogènes (Encadré 1.5). De manière alternative, Papaïx et al. (2014), dont j'ai parlé dans la partie précédente, ont utilisé un modèle stochastique pour rendre compte du processus aléatoire de l'épidémie, via la transition entre états pour les plantes (saines, infectées mais non infectieuses, infectées et infectieuses, et 'retirées', c'est-

à-dire qui ont déjà été infectées et infectieuses et ne peuvent plus l'être). Lo Iacono et al. (2013) étudient l'effet des stochasticités démographique et environnementale sur les populations d'agents pathogènes et sur les stratégies de gestion de la résistance des plantes. Le paysage qu'ils étudient comprend deux variétés de plantes, une sensible et une résistante, et deux variants pathogènes, un avirulent et un virulent. En comparant leur modèle à un modèle déterministe équivalent sans stochasticité démographique, les auteurs ont montré que, sans prendre en compte la stochasticité démographique, la proportion de plantes résistantes dans le paysage avait un effet négligeable sur la durabilité de la résistance alors que la stochasticité démographique rendait cet effet important. Avec la version déterministe, augmenter la proportion de plantes résistantes conduit à une diminution des pertes de rendements, mais la résistance est plus vite contournée, il y a donc un compromis à trouver entre capacité d'évolution de l'agent pathogène et gains de rendement. La stochasticité démographique annule cet effet de compromis. Ils ont également montré que des fluctuations importantes du nombre d'hôtes infectés, à cause par exemple de l'application d'un fongicide, augmentait les chances d'extinction de l'agent pathogène dans la version stochastique du modèle. Ces faibles nombres d'hôtes infectés sont retrouvés au début des épidémies. Cette condition se retrouve forcément au début de chaque saison de culture dans les systèmes saisonniers. Les auteurs insistent sur l'utilisation de modèles stochastiques pour de tels systèmes. Prendre en compte la stochasticité des processus a un effet important sur les conclusions que l'on peut tirer sur le fonctionnement des épidémies dans les cultures et l'effet des différentes stratégies de contrôle.

2.4.2 Stratégies de gestion du paysage

Comme évoqué dans le chapitre 1, différentes stratégies de déploiement de variétés résistantes ont été étudiées, telles que l'utilisation de variétés multi-résistantes (pyramidage de plusieurs gènes de résistance), le mélange de variétés ou la rotation dans les cultures (McDonald and Linde, 2002a; Mundt, 2002). Différentes études, expérimentales et théoriques, ont pu prouver leur efficacité et montrer dans quelles conditions ces stratégies donnaient les meilleurs résultats en termes de gains de rendement et de durabilité de la résistance des plantes.

Mélange de variétés

Il y a un peu plus de 60 ans, les premiers travaux sur l'intérêt de la diversité génétique dans les plantes cultivées, par l'utilisation de variétés multi-résistantes, ont été publiés par Jensen (1952). Depuis, de nombreux travaux se sont intéressés au mélange de variétés, notamment

d'une variété sensible et d'une variété résistante à un agent pathogène (e.g. Fabre et al., 2012b; Mundt, 2002; Ohtsuki and Sasaki, 2006; Pink and Puddephat, 1999). Dans sa revue, Mundt (2002) détaille l'effet de quatre facteurs sur le bénéfice d'utiliser un mélange de variétés pour le contrôle d'une maladie dans un paysage agricole. Le premier facteur est la proportion d'autoinfections⁵ par rapport aux alloinfections⁶, influencée par les caractéristiques de l'agent pathogène et de l'hôte. Il est attendu que l'efficacité des mélanges variétaux pour le contrôle d'une maladie décroisse avec une proportion croissante d'autoinfections (e.g. Barrett, 1980). Deuxièmement, des modèles mathématiques ont montré que l'efficacité des mélanges diminuait avec la profondeur du gradient de dispersion de l'agent pathogène (e.g. Fitt and McCartney, 1986). Troisièmement, des études numériques ont montré que la croissance des lésions peut grandement diminuer l'efficacité des mélanges (e.g. Lannou et al., 1994). En effet, la croissance d'une infection peut contribuer à l'autoinfection, mais également au développement de l'épidémie. Enfin, des études numériques et expérimentales au champ ont montré que l'efficacité des mélanges variétaux diminuait lorsque le degré d'aggrégation des variétés était changé et altérait l'unité d'aire génotypique⁷ (e.g. Mundt and Browning, 1985). Plus précisément, le contrôle de la maladie est meilleur lorsque les variétés sont mélangées de manière plus fine (petite unité d'aire génotypique), plutôt que d'alterner des parcelles contenant chacune une variété uniquement. Cela a notamment été montré pour les cultures de riz au Japon, le mélange de variété au sein d'une colline donnant de meilleurs résultats qu'entre collines pour le contrôle de la pyriculariose (Koizumi, 2001).

Plus généralement, dans les mélanges de deux variétés, une sensible et l'autre résistante à un agent pathogène, la variété sensible peut introduire un effet de dilution de l'inoculum dans la population d'agents pathogènes (Mundt, 2002). En effet, les variants pathogènes adaptés sont généralement contre-sélectionnés dans les plantes sensibles à cause du coût de fitness des mutations de virulence, ce qui maintient les variants non adaptés dans le paysage et en retour les infections d'une plante sensible vers une plante résistante seront souvent infructueuses (Fabre et al., 2012b).

Chez les virus et bactéries des plantes transmis par vecteur, Mundt (2002) précise que les mélanges variétaux peuvent avoir des conséquences diverses sur le contrôle d'une maladie, car ils ont un impact sur le comportement des vecteurs et la transmission du virus. Par exemple, le meilleur contrôle de la jaunisse nanisante de l'avoine grâce au mélange variétal peut être dû au fait que les pucerons vecteurs montrent des taux de mouvement plus élevés et des temps

5. L'individu hôte donneur infecté est le même que l'individu hôte receveur.

6. L'hôte donneur est un individu différent du receveur.

7. Surface occupée par une variété unique.

d'alimentation plus courts par rapport à une monoculture. En effet, pour cette maladie le puceron doit s'alimenter plusieurs heures pour transmettre le virus (mode de transmission circulant) (Power, 1991). Par opposition, un mélange de maïs et de haricots a considérablement réduit la densité de cicadelles (vecteurs), mais pas l'incidence de la maladie du rabougrissement du maïs. Dans ce cas, l'augmentation du mouvement des vecteurs et le raccourcissement des temps d'alimentation sur chaque plante a augmenté le nombre de transmissions car l'agent pathogène pouvait être transmis après des temps d'alimentation très courts (Power, 1987).

Pour terminer, je décrirai à titre d'illustration quelques résultats d'une étude épidémiologique à laquelle j'ai contribué, sur l'effet de différents facteurs sur la réduction des dégâts relatifs, c'est-à-dire les pertes de rendement dans un paysage dans lequel est déployée une variété résistante à un virus, en mélange avec une variété sensible, par rapport à un paysage avec la variété sensible uniquement (Fabre et al., 2012b). Nous avons pu mettre en évidence que le facteur le plus important pour les dégâts relatifs, et donc pour le contrôle de la maladie était l'intensité des épidémies, pouvant être vue comme le taux moyen de transmission de la maladie dans un paysage composé uniquement de plantes sensibles, représentant l'activité moyenne des vecteurs et le taux de réussite de l'infection suite à l'alimentation sur une plante saine. Une intensité d'épidémie plus faible permettait généralement une plus forte diminution des dégâts relatifs. Le deuxième facteur le plus important était le choix du gène majeur de résistance, englobant le nombre de mutations nécessaires au contournement et le coût de fitness associé à chaque mutation dans les plantes sensibles. Un gène majeur de résistance nécessitant plus de mutations pour être contournée avec des coûts de fitness plus élevés conduisait généralement à une diminution plus importante des dégâts relatifs, car alors la résistance était plus difficilement contournable. Les deux facteurs suivants étaient la proportion de plantes résistantes dans le paysage et l'importance relative des trois voies possibles d'infection : depuis le réservoir (plantes sauvages) vers les parcelles cultivées, entre parcelles et au sein d'une parcelle. Lorsque les infections se faisaient principalement entre parcelles, des proportions intermédiaires de plantes résistantes étaient préférables pour un meilleur contrôle de la maladie. Par contre, lorsque les infections se faisaient principalement depuis le réservoir, des proportions élevées de plantes résistantes étaient optimales, allant jusqu'à 100% de plantes résistantes. Au final, cette étude montre que le mélange de variétés peut constituer une bonne stratégie pour le contrôle d'une maladie virale dans les plantes, mais l'efficacité de cette stratégie dépend d'autres facteurs, ici en particulier de l'importance relative des différentes voies d'infection, mais aussi de l'intensité des épidémies et des caractéristiques du gène majeur de résistance.

Rotations dans les cultures

Un autre stratégie de gestion de la résistance des plantes consiste à procéder à des rotations dans les cultures. Notamment, dans une étude récente à laquelle j'ai également contribué, nous avons cherché à optimiser les proportions de plantes résistantes dans le paysage en autorisant des rotations tous les 3 ans, sur 15 ans de cultures, et nous avons comparé ces stratégies de rotation des cultures avec des stratégies sans rotation (uniquement un mélange de variétés) (Fabre et al., 2015). Dans cet article comme dans l'article Fabre et al. (2012b) décrit précédemment, le paysage était composé de trois voies d'infection : depuis le réservoir vers les parcelles cultivées, entre parcelles et au sein d'une parcelle. Nous avons ainsi pu montrer que les rotations peuvent être particulièrement bénéfiques par rapport au déploiement d'une proportion constante de plantes résistantes (i) lorsque les infections se déroulent majoritairement au sein de chaque parcelle (faible distance de dispersion), ou (ii) lorsqu'elles se déroulent principalement depuis le réservoir et que les gènes de résistance demandent une seule mutation pour être contournés. De plus, de telles stratégies de rotations peuvent permettre de diminuer les pertes de rendement tout en veillant à ce que les variants adaptés restent en-dessous d'un certain seuil, augmentant ainsi la durabilité de la résistance.

Dans les domaines de la santé publique et de l'agronomie, REX Consortium (2013) ont mis en avant l'intérêt des combinaisons entre mélange de molécules et leurs rotations dans le temps pour la gestion de la résistance conférée par les insecticides ou antibiotiques aux ravageurs et aux agents pathogènes, après avoir analysé un ensemble de 29 études théoriques sur le sujet. Ces études comparaient cette stratégie double (mélange et rotation) à des stratégies plus simples, consistant (i) en un mélange de molécules constant dans le temps, ou (ii) en une alternance temporelle entre différentes molécules utilisées une par une séparément, soit de manière cyclique, soit en changeant de molécule lorsque celle en cours d'utilisation était contournée. REX Consortium (2013) ont conclu que les stratégies combinant mélange et rotations étaient au moins aussi performantes, voire plus performantes que les autres stratégies pour retarder le contournement d'une résistance dans plus de 80% des comparaisons.

2.4.3 Épidémiologie évolutive

L'épidémiologie évolutive fait appel à l'écologie évolutive ainsi qu'à la théorie de l'épidémiologie (Galvani, 2003). Ainsi, elle s'intéresse à des caractéristiques aussi diverses que la structuration spatiale des populations d'hôtes et d'agents pathogènes, les dynamiques intra-hôtes, l'hétérogénéité des hôtes, leur longévité et leur taille de population (Galvani, 2003). Pour

les agents pathogènes des plantes comme pour ceux des animaux, l'épidémiologie évolutive doit intégrer les différents niveaux auxquels les agents pathogènes sont soumis aux forces évolutives, depuis le niveau intra-hôte jusqu'au niveau inter-hôtes (Galvani, 2003; Mideo et al., 2008; Park et al., 2013). Notamment, Galvani (2003) souligne l'importance d'intégrer tous ces aspects aux différents niveaux concernés pour pouvoir mieux appréhender les questions de santé publique ou de gestion des cultures face aux agents pathogènes. La prise en compte des deux niveaux se fait généralement par l'utilisation de modèles emboîtés, incorporant un modèle représentant les dynamiques des populations d'agents pathogènes au niveau intra-hôte dans un modèle représentant les dynamiques épidémiques au niveau d'une population d'hôtes (Mideo et al., 2008). Par exemple, Park et al. (2013) ont développé une méthode constituée d'un modèle mécaniste représentant la sélection intra-hôte, emboîté dans un modèle stochastique au niveau inter-hôte, permettant d'étudier l'émergence de variants pathogènes adaptés face à un nouvel environnement (par exemple une nouvelle variété d'hôte résistante ou un traitement médicamenteux). Ils ont montré que des corrélations positives dans les fitness entre les deux niveaux pouvaient faciliter l'émergence de variants pathogènes adaptés, alors que des sélections contraires entre niveaux pouvaient conduire à des extinctions de la population d'agents pathogènes.

Les études d'épidémiologie évolutive s'intéressent souvent à l'évolution des traits d'histoire de vie des agents pathogènes et aux éventuels compromis évolutifs pouvant exister entre eux. Les traits de vie les plus couramment étudiés sont l'agressivité (appelée virulence en biologie évolutive, mais je garderai le terme d'agressivité par souci d'homogénéité dans l'ensemble de ce document) et la transmission. Pour reprendre l'idée de l'importance des différents niveaux auxquels évoluent les agents pathogènes, les interactions entre différents variants pathogènes au niveau intra-hôte peuvent avoir une influence importante sur la sélection de variants plus agressifs (Galvani, 2003). C'est par exemple le cas pour la compétition entre variants du *Human immunodeficiency virus* (genre *Lentivirus*) (HIV), dont la croissante agressivité a été associée à l'émergence du syndrome d'immunodéficience acquise (SIDA). Les modèles d'évolution de l'agressivité prenant en compte les deux niveaux intra- et inter-hôtes montrent que la réduction de la transmission ainsi que la réduction du nombre d'hôtes infectés par plusieurs variants pathogènes peut sélectionner des variants pathogènes montrant moins d'agressivité (Galvani, 2003).

Concernant l'hétérogénéité des hôtes (mélange de variétés), il a été montré que si un variant pathogène est bien adapté à un génotype d'hôte résistant et mal adapté aux autres génotypes présents dans la population d'hôte, alors l'hétérogénéité des hôtes va sélectionner

une agressivité moins forte (Regoes et al., 2000), ainsi qu'une hétérogénéité dans l'agressivité (Galvani, 2003; Ganusov et al., 2002).

Pour faire le lien avec la notion de saisonnalité évoquée plus haut, van den Berg et al. (2011) ont étudié l'effet de l'absence périodique des hôtes sur le taux de transmission d'agents pathogènes. Ils ont montré qu'une plus longue période d'absence périodique des hôtes pouvait entraîner la sélection d'agents pathogènes avec des taux de transmission soit plus forts, soit plus faibles, selon la nature de deux compromis évolutifs étudiés, respectivement transmission - agressivité ou transmission - survie en l'absence d'hôtes. Dans leur revue, Zhan et al. (2015) évoquent cette absence périodique de l'hôte comme caractéristique des paysages agricoles favorisant la dérive génétique, conduisant à une diminution de la variabilité génétique des populations d'agents pathogènes, et donc permettant d'augmenter la durabilité d'une résistance.

De manière très intéressante, May and Anderson (1983) ont pu formuler le taux de reproduction de base R_0 en fonction de l'agressivité et de la transmissibilité des agents pathogènes :

$$R_0 = \frac{\beta_{N\alpha}}{\alpha + \mu + \nu_\alpha}.$$

Dans cette formule, α est le taux d'agressivité ou de mortalité liée à l'infection, $\beta_{N\alpha}$ est le taux de transmission de la maladie (fonction de la taille de la population d'hôtes N et de α), μ est le taux moyen de mortalité des hôtes sains et ν est le taux de rétablissement suite à l'infection (fonction de α). De manière similaire au R_0 défini précédemment, il correspond au nombre moyen d'infections secondaires dues à un hôte infecté. Cette formule de R_0 repose sur l'hypothèse que les différents paramètres qui la composent ne dépendent pas de la densité des agents pathogènes (Galvani, 2003). On voit qu'une forte agressivité de l'agent pathogène diminue le R_0 , alors qu'un taux de transmission élevé l'augmente. D'une certaine manière, on retrouve l'idée d'un compromis évolutif entre transmission et agressivité pour la persistance de la maladie.

Ainsi, nous avons pu voir un ensemble d'études sur des thématiques diverses rattachées pour la plupart à l'épidémiologie végétale. Plus particulièrement nous avons pu voir l'intérêt des études théoriques en lien avec la durabilité de la résistance des plantes aux agents pathogènes pour un meilleur contrôle des maladies dues à ces derniers. Ces études théoriques permettent de faire des prédictions au niveau d'une population d'hôtes pour une large gamme de valeurs de paramètres d'intérêt, ce qui est difficile à réaliser en conditions naturelles. Des études biologiques en conditions naturelles ont été conduites, mais elles ne peuvent pas se permettre

d'explorer l'ensemble des cas étudiés théoriquement. Les études biologiques sont cependant très importantes, car elles permettent d'évaluer leur cohésion avec les études théoriques, voire de faire émerger une théorie plus approfondie et d'en améliorer le pouvoir prédictif.

Concernant leur lien avec les forces évolutives agissant sur les populations d'agents pathogènes, l'ensemble de ces études théoriques en épidémiologie végétale s'intéresse principalement à la sélection. À ma connaissance, aucune étude ne s'est intéressée à l'effet de la dérive génétique agissant sur les populations d'agents pathogènes sur la durabilité de la résistance des plantes. Pourtant, l'absence périodique des hôtes typique des systèmes agricoles en climat tempéré, ainsi que les nombreux goulets d'étranglement subis plus particulièrement par les virus dans leurs plantes hôtes sont des facteurs importants augmentant l'intensité de la dérive génétique. Comme discuté par Zhan et al. (2015), mettre en place des outils permettant d'augmenter la dérive génétique pourrait permettre un meilleur contrôle des maladies des plantes sur le long terme, et des gains de rendement plus importants. C'est précisément le thème de l'article présenté dans le chapitre suivant, dans lequel je me suis intéressée à l'effet de la dérive génétique agissant sur les populations virales sur les gains de rendement additionnels pouvant être atteints et sur la durabilité d'une résistance qualitative déployée en mélange avec une variété sensible dans une parcelle agricole.

Structure de la thèse

Avant de passer au chapitre 3, je vais expliquer ici l'enchaînement des trois articles de cette thèse. Dans l'introduction, nous avons vu les deux types de résistance des plantes, résistance qualitative et résistance quantitative, et plus particulièrement nous avons pu voir l'intérêt de combiner ces deux types de résistance pour augmenter la durabilité de la résistance qualitative (Brun et al., 2010; Fournet et al., 2013; Palloix et al., 2009; Quenouille et al., 2013, 2014). Nous avons également passé en revue les forces évolutives agissant sur les agents pathogènes dans leurs plantes hôtes ainsi que lors de la transmission d'hôte à hôte, et nous avons discuté de l'intérêt d'intégrer les connaissances sur les principes éco-évolutifs pour une meilleure gestion des agents pathogènes dans les cultures (Brown, 2015; García-Arenal et al., 2001, 2003; McDonald and Linde, 2002a; Zhan et al., 2015).

Dans la suite de ce manuscrit, je commence par étudier l'intérêt de la dérive génétique au niveau des épidémies via une approche de modélisation. Pour cela, j'ai analysé l'effet d'une plus forte intensité de dérive génétique lors de l'infection des plantes et lors de la transmission entre plantes pour l'augmentation des gains de rendement dans les cultures (article 1). Cet article s'inscrit dans la continuité de ceux de Palloix et al. (2009) et Quenouille et al. (2013, 2014) sur l'effet de la présence de résistances quantitatives dans le fond génétique du piment sur la fréquence de contournement du gène majeur de résistance *pvr2³* vis-à-vis du PVY. Notamment, Quenouille et al. (2013) ont émis l'hypothèse que la plus faible fréquence de contournement de ce gène majeur en présence de résistances quantitatives pouvait être due à une réduction des tailles efficaces de populations virales lors de l'infection d'une plante, conduisant à une plus forte dérive génétique. J'ai cherché à évaluer cette hypothèse au niveau d'une population d'hôtes, grâce à un modèle épidémiologique stochastique mimant des goulets d'étranglement plus ou moins étroits subis par les populations virales, depuis la transmission de plante à plante jusqu'au début de l'infection systémique d'une plante. En effet, ces étapes peuvent être associées à des petites tailles de populations virales, et donc potentiellement à de fortes intensités de dérive génétique. Grâce à ce modèle, j'ai pu montrer l'intérêt de la dérive génétique pour diminuer les pertes

de rendement, et donc pour une meilleure gestion de la résistance qualitative des plantes. Vu son intérêt, je me suis alors demandée s'il était possible de contrôler cette dérive génétique grâce au génome des plantes pour pouvoir effectivement l'utiliser dans les cultures. J'ai donc essayé d'identifier des variations significatives entre génotypes de plante dans les intensités de la dérive génétique, mais également de la sélection subies par les populations virales lors de l'infection des plantes. Cette démarche va plus loin que celle de Quenouille et al. (2013) qui ont mesuré l'accumulation virale dans les plantes et les fréquences de contournement du gène majeur *pvr2³*, puis ont formulé des hypothèses sur le rôle potentiel de la sélection et de la dérive génétique dans ces résultats, sans avoir mesuré l'intensité de ces forces évolutives. Je me suis donc attachée à estimer les tailles efficaces de populations virales ainsi que les coefficients de sélection de variants viraux au cours de l'infection de différents génotypes de piment (article 2, chapitre 4). Pour cela, je me suis basée sur un modèle mécanistico-statistique proposé par Fabre et al. (2012a) que j'ai amélioré puis validé grâce à des simulations numériques du modèle de Wright-Fisher avec des valeurs connues de N_e et s . J'ai ainsi pu montrer une forte variabilité dans les valeurs de N_e et s caractérisant les populations virales d'un génotype de plante à un autre, ainsi qu'un contrôle génétique de ces deux caractères par les plantes. Enfin, dans le troisième article (chapitre 5), j'ai effectué une analyse statistique globale des estimations de N_e et s du chapitre 4 et des estimations des fréquences de contournement du gène majeur *pvr2³* et d'accumulation virale de Quenouille et al. (2014). Cette analyse m'a permis d'étudier l'effet de la dérive génétique, de la sélection et de l'accumulation virale sur la durabilité de la résistance qualitative, à partir de données expérimentales et non plus par une approche théorique. J'ai ainsi pu mettre en évidence un lien significatif entre la dérive génétique aux stades précoces de l'infection (à l'inoculation et dans la feuille inoculée) et la fréquence de contournement du gène *pvr2³*. Ces derniers résultats sont complémentaires de ceux de l'article 1 puisqu'ils portent sur le contournement d'une résistance qualitative, mais cette fois-ci au niveau d'une plante hôte et non pas au niveau d'une population d'hôtes.

Pour résumer, je vais vous présenter dans le chapitre suivant un modèle mathématique épidémiologique pour l'étude de l'effet de la dérive génétique intra-hôte et lors de la transmission inter-hôtes sur la diminution des pertes de rendement au niveau d'une population d'hôtes. Dans le chapitre 4, je présenterai une méthode d'estimation de N_e et s caractérisant les populations virales lors de l'infection d'un hôte, ainsi que les estimations de ces deux paramètres sur quinze génotypes de piment. Enfin, dans le chapitre 5, je présenterai les analyses statistiques sur l'effet de N_e , s et de l'accumulation virale sur les fréquences de contournement du gène majeur *pvr2³*.

Chapitre 3

Modulating genetic drift exerted on virus populations to increase the durability of plant resistance

List of authors : Elsa Rousseau^{1,2,3}, Mélanie Bonneault¹, Frédéric Fabre⁴, Benoit Moury³, Ludovic Mailleret^{1,2}, Frédéric Grognard¹

¹ Biocore Team, INRIA, F-06902 Sophia Antipolis, France

² INRA, Univ. Nice Sophia Antipolis, CNRS, UMR 1355-7254 Institut Sophia Agrobiotech, 06900 Sophia Antipolis, France

³ UR 407 PV, INRA, F-84140 Montfavet, France

⁴ UMR 1065 SAVE, INRA, F-33882 Villenave d'Ornon, France

Abstract

Background

Plant qualitative resistances against viruses are a natural exhaustible resource, which can be impaired by the emergence of resistance-breaking (RB) variants. Modeling approaches can help to determine optimal strategies for resistance durability and crop damage reduction by a rational deployment of resistance in agroecosystems. Here, we propose an innovative tool, based on the introduction of quantitative resistance in the genome of plants to decrease the bottleneck sizes experienced by viruses from their transmission between hosts to the infection within a host, and thereby to increase the strength of genetic drift. Genetic drift is expected to slow down the virus adaptation to plant resistance. To study the effect of genetic drift on yield increase, we developed a stochastic plant epidemic model relying on continuous time Markov chains processes.

Results

Overall, genetic drift did not bring substantial improvements when the qualitative resistance was already highly efficient alone or when the RB variant was already present at a high frequency in the agroecosystem. In all the other cases, genetic drift was beneficial, and could provide up to 95 percentage points additional yield compared to strategies deploying qualitative resistance without bottleneck quantitative resistance (35 percentage points in average).

Conclusion

Our model showed the benefit of genetic drift to increase yield and potentially delay the emergence of RB variants. As the intensity of genetic drift experienced by virus populations is a trait that can be controlled genetically by host plants, our results could promote its use in breeding programs and the application of resistance deployment strategies by farmers.

Keywords : qualitative resistance durability, quantitative resistance, population bottleneck, genetic drift, yield increase, stochastic modeling.

3.1 Introduction

Integrating the principles of population genetics into the management of plant genetic resistance to pathogens in agricultural ecosystems should provide a powerful way to achieve sustainable disease resistance (McDonald and Linde, 2002a; Zhan et al., 2014, 2015). This will help to face up to the challenge of food security while decreasing the environmental impact of agriculture (Mundt, 2014; Tilman, 1999). The evidence of that lies in natural ecosystems, where the coevolutionary processes between wild plants and their associated pathogens limit the spatial extent and temporal duration of disease epidemics (Zhan et al., 2015). In modern agricultural ecosystems, (i) intensified monocultures of superior yield and quality varieties, inducing a high selection pressure on the pathogens, and (ii) imposing cultivated plant varieties in the landscape, preventing the natural coevolution between plants and pathogens, have disrupted these processes, leading to increased risks of widespread plant diseases and thus yield losses (Zhan et al., 2015). It has been shown that adopting strategies like the use of multiline cultivars or cultivar mixtures can decrease the spread of epidemics (Fabre et al., 2012b; Mundt, 2002; Wolfe, 1985). More generally, maintaining higher degrees of cultivar heterogeneity over time and space is associated with slower pathogen adaptation (Fabre et al., 2015; Gilligan, 2008; REX Consortium, 2013).

Plant disease resistance often results from a molecular relationship governed by the gene-for-gene concept (Flor, 1971). Plant qualitative resistances, i.e. resistances that almost totally prevent any plant infection, follow this pattern. In this way, the interaction between the product of the resistance gene of the plant (with at least two allelic forms : 'resistant' (R) and 'susceptible' (S)) and the avirulence gene of the pathogen (with at least two allelic forms : 'wild-type' (WT) and 'resistance-breaking' (RB)) determines the ability of the pathogen to infect a plant. With the two allelic forms in both the host and the pathogen, four types of interactions are possible, the only one preventing plant infection (incompatible interaction) being the WT variant/R plant interaction. In the three other cases, the interaction is compatible and the pathogen can infect the plant.

The gene-for-gene system, however, does not often provide durable resistance (Brown, 2015), especially for fungal and bacterial pathogens (Parlevliet, 2002), but also for viruses (García-Arenal and McDonald, 2003). For the latter, only one or two mutations in the pathogen genome may be sufficient for the pathogen to break down the resistance (RB allelic form) (Brown and Hovmøller, 2002; Harrison, 2002; Lecoq et al., 2004). Once the RB variant has appeared, it has a disproportionate selective advantage to settle in a hitherto resistant cultivar, and indeed

in all fields with the same cultivar (or even other cultivars with the same resistance gene) (Brown, 2015). The time necessary for a pathogen to break down a plant resistance is highly variable, especially for viruses (Kang et al., 2005), showing both highly durable resistances, as the *Tm2* gene in tomato or the *N*-gene in tobacco against the *Tobacco mosaic virus* (TMV, genus *Tobamovirus*), and poorly durable resistances, as the *Tm2* gene in tomato against TMV (Parlevliet, 2002).

A second type of plant genetic resistance is quantitative resistance, controlled by the genome of the plant. Pyramiding a qualitative resistance with quantitative resistance is a promising tool to increase the durability of the qualitative resistance (Brun et al., 2010; Palloix et al., 2009). Notably, a recent laboratory study showed that pyramiding the *pvr2³* qualitative resistance with quantitative resistance in the genome of pepper plants improved the durability of *pvr2³* against *Potato virus Y* (PVY) (Palloix et al., 2009). Three mechanisms were later proposed and validated to explain this result, observed in laboratory condition at the host plant scale : (i) a higher efficiency of the qualitative resistance provided by quantitative resistance thanks to the reduction of pathogen multiplication, (ii) more complex mutational pathways for the pathogen to break down the resistance, and (iii) a reduction in the fixation probability of beneficial mutations for resistance breakdown (Quenouille et al., 2013). In fact, population genetics theory teaches us that two main evolutionary forces can modulate the fixation probability of a beneficial mutation : selection and genetic drift (Kimura, 1970; Patwa and Wahl, 2008; Sniegowski and Gerrish, 2010). Selection is a deterministic force favoring the fittest variants and is generally evaluated with the selection coefficient between pathogen variants. Genetic drift produces random fluctuations in the evolution of variant frequencies, eventually purging variants regardless of whether they are beneficial, neutral or harmful (Zhan et al., 2015). It is generally evaluated with the effective population size (Charlesworth, 2009). For a fixed value of the selection coefficient of a beneficial mutation, the fixation of this mutation is generally slowed down when genetic drift is strong, i.e. small effective population size (Patwa and Wahl, 2008), because the important stochastic fluctuations in variant frequencies can distort the direction of evolution given by selection (Gutiérrez et al., 2012a).

It has been shown that the intensity of genetic drift can be genetically controlled by the host plant (Rousseau et al., Submitted), with effective population sizes ranging from 25 to 1059 in the inoculated leaves and from 28 to 390 at the onset of systemic infection, i.e. colonization of apical leaves, depending on the plant genotype. More generally, small effective population sizes can be observed during plant infection by viruses because population bottlenecks are rather common for those pathogens (Gutiérrez et al., 2012a; Zwart and Elena, 2015). During within-

host progression, i.e. colonization of new leaves of a plant, narrow or larger bottlenecks may be present, depending on the plant - virus pair, leading to effective population sizes in the range 1-5.6 in the former case (Fabre et al., 2012a; French and Stenger, 2003; Sacristán et al., 2003; Zwart et al., 2011) or in the range 1.2-131 in the latter case (Gutiérrez et al., 2012b; Zwart et al., 2011). During host-to-host transmission by insect vectors, effective population sizes may be reduced to 0.5-3.2 virus individuals or virus infectious units in average (Betancourt et al., 2008; Moury et al., 2007). Quantitative resistance modulating the virus effective population size at mechanical inoculation have already been identified (Tamisier et al., 2016), and we believe it possible to identify such quantitative resistance at other stages of plant infection.

Disease management strategies that increase the strength of genetic drift by a reduction of the pathogen effective population size will help control epidemics by limiting the genetic diversity in the pathogen population and hence slow down pathogen evolution (McDonald and Linde, 2002a; Zhan et al., 2015). We are interested in the use of bottleneck quantitative resistance, decreasing pathogen effective population size, to increase the durability of a qualitative resistance. Specifically, in this paper, we present a stochastic plant epidemic model coupling epidemiology and population genetics to study the effect of random genetic drift, controlled by quantitative resistance, on qualitative resistance durability. Our model is based on classical healthy (susceptible) - infected (SI) deterministic model (Brauer et al., 2008), representative of plant virus epidemics. The term 'susceptible' is usually employed in epidemiology to refer to a healthy plant. We use instead the term 'healthy', since the term 'susceptible' is used in phytopathology to qualify plants that carry no resistance gene. We focus on viruses, but the proposed framework should also be relevant for fungal or bacterial pathogens. The aim of our study was to compare the yield benefits provided by a resistant cultivar carrying both qualitative and quantitative resistances (named polygenic resistance) with those provided by a resistant cultivar carrying only the qualitative resistance in a plant susceptible genetic background (named monogenic resistance). This allows to predict the added value of pyramiding quantitative resistance controlling virus effective population size, and thus the intensity of genetic drift exerted on virus populations, to improve the long-term efficiency of a qualitative resistance. We investigated the interactions between agroecosystems features (intensity of epidemics, resistance deployment strategy), the characteristics of the qualitative resistance genes (through their effects on the fitness cost imposed to the RB variants) and the characteristics quantitative resistance genes (through their effects on the effective population size of the pathogen population) on yield benefits. We found that the knowledge of the factors representative

of those characteristics is essential in order to identify the best strategy for yield benefits.

3.2 Materials and methods

3.2.1 Model overview

The model merges epidemic processes at the field scale, involving viruses, plants and vectors, and virus population genetics processes, and takes its roots in a model developed by Fabre et al. (2012b). It describes, during a cropping season lasting n_d days, the dynamics of epidemics in a field composed of susceptible (S) and resistant (R) plants (i.e. plants carrying a qualitative resistance). The field consists of a constant number of plants, N^p , amongst which a proportion φ are R plants. Hence the number of S plants is $N^S = (1 - \varphi)N^p$ and the number of R plants is $N^R = \varphi N^p$. Two virus variants are considered, namely the 'wild-type' and the 'resistance-breaking' (RB) variants. The ability of the virus to infect a plant is assumed to be governed by the gene-for-gene interaction system (Flor, 1971). Thus, only the RB variant can infect R plants, and both variants can infect S plants. Bottlenecks undergone by virus populations are considered from the host-to-host transmission step until the onset of systemic infection. They include the host-to-host transmission process by vectors, and the following within-host progression processes : colonization of the inoculated leaves (cell-to-cell movement) and passage in the vascular system to reach apical leaves (Gutiérrez et al., 2012a). During these steps, we assume that genetic drift is the main evolutionary force at stake and the other forces - selection and mutation - can be neglected. All bottleneck sizes and the transitions between bottlenecks are summarized into a unique effective population size accounting for the global effect of all bottlenecks, denoted N_e^R in R plants and N_e^S in S plants. Values of N_e^R could be reduced thanks to the presence of quantitative resistance in R plants (i.e. plants pyramiding qualitative and quantitative resistances). On the opposite, S plants never carried quantitative resistance reducing N_e^S value. After the bottlenecks have been crossed, i.e. after the onset of systemic infection, genetic drift is neglected. We assume that virus populations reach instantaneously their mutation - selection - genetic drift equilibrium within the host plants, with the frequency of the RB variant in S plants being denoted f_{RB} . The f_{RB} parameter is characteristic of the choice of the resistance gene, depending on the number of mutations required for resistance breakdown and on the associated fitness costs (Fabre et al., 2012b). The main variables and parameters of the model are described in table 3.1.

Hereafter, we describe the model in its deterministic form, for a field composed only

of S plants, without quantitative resistance. This case defines the baseline for the various epidemiological contexts studied, before deployment of R plants. It is also used to attribute meaningful values to a parameter representative of the intensity of epidemics in a reference field before deployment of R plants. Then, we briefly describe this deterministic model with R plants and finally move to its stochastic form, used for all simulations of epidemics after the deployment of R plants. More precisely, for each epidemiological context, we will investigate the effect of the introduction of monogenic resistant plants, i.e. without quantitative resistance, and, further, of the introduction of polygenic resistant plants, with quantitative resistance reducing virus effective population size, on yield benefit.

TABLE 3.1 – Description of the parameters and state variables of the model, and of the factorial design for numerical simulations.

Parameters	Designation	Unit	Range or reference value
Ω_{int}	Epidemic intensity before deployment of R plants	Unitless	[0.1, 0.9]
n_d	Duration of the cropping season	Day	120
N^P	Number of plants in the field	Plant	10^3
φ	Proportion of R plants	Unitless	[0.05, 0.95]
f_{RB}	Choice of R gene : frequency of the RB variant in a S plant	Unitless	$[10^{-8}, 0.5]$
N_e^R	Virus effective population size in R plants	Virus or infectious unit	Small : [1, 100] Large : 10^4
N_e^S	Virus effective population size in S plants	Virus or infectious unit	10^4
n_{iter}	Number of simulation iterations for each set of parameter values	Unitless	500
State Variables			
$I^S(t)$	Number of infected S plants at time t	Plant	
$I^R(t)$	Number of infected R plants at time t	Plant	

3.2.2 Description of the deterministic model

In a fully susceptible field ($\varphi = 0$)

In this situation, the host carries no resistance at all (qualitative or quantitative) and the virus effective population size N_e^S accounting for bottlenecks from host-to-host transmission to within-host progression is very large (10^4). The dynamics of I^S , the number of infected S plants across the n_d days of the cropping season was assumed to follow a healthy - infected

type ordinary differential equation (ODE) (Brauer et al., 2008), as :

$$\frac{dI^S}{dt} = \beta_0 \frac{I^S(N^S - I^S)}{N^S} \left(1 - e^{-N_e^S}\right). \quad (3.1)$$

Only one infected S plant is considered at the beginning of the cropping season, i.e. $I^S(0) = 1$. In equation 3.1, β_0 is the basic contact rate from an infected plant to a healthy one through insect vectors. The last term of this equation accounts for genetic drift in S plants thanks to the virus effective population size N_e^S . We suppose that S plants get infected if at least one virus particle (or infectious unit) passes through all bottlenecks from the entry in the host until its systemic infection. Assuming that the number of virus particles surviving all bottlenecks results from a Poisson distribution of mean N_e^S , it follows that the probability that at least one virus particle survives all bottlenecks is $1 - e^{-N_e^S}$. In fact, as the S plant carries no quantitative resistance, i.e. $N_e^S = 10^4$, we have $1 - e^{-N_e^S} \simeq 1$. Hence the effect of genetic drift in this simple model is negligible.

The analytic integration of the proportion of infected S plants $p_i^S(t) = I^S(t)/N^S$ over n_d days through equation 3.1 provides the area under the disease progress curve (AUDPC) in the field, $A_0 = \int_0^{n_d} p_i^S(t) dt = \frac{1}{\beta_0} \ln \left(1 + p_i^S(0) (e^{\beta_0 n_d} - 1)\right)$. Following Fabre et al. (2012b), we use the AUDPC to attribute meaningful values to the epidemic parameter β_0 . For this purpose, we introduce a new parameter, the intensity of epidemics Ω_{int} , giving the average proportion of plants infected along a cropping season in a fully susceptible field. Thereby we have the relationship $\Omega_{int} = A_0/n_d$, from which we can deduce values of β_0 for various epidemic intensities Ω_{int} (Fabre et al., 2012b).

Deployment of resistant plants

Let us now consider that R plants are also present in the field, so that I^R denotes the number of infected R plants during the cropping season. The model describing the epidemics then reads as :

$$\left\{ \begin{array}{l} \frac{dI^S}{dt} = \left(\beta_0 \frac{I^S(N^S - I^S)}{N^S} + \beta_0 \frac{I^R(N^S - I^S)}{N^S} \right) \left(1 - e^{-N_e^S}\right) \end{array} \right. \quad (3.2)$$

$$\left\{ \begin{array}{l} \frac{dI^R}{dt} = \beta_0 \frac{I^R(N^R - I^R)}{N^R} \left(1 - e^{-N_e^R}\right) + \beta_0 \frac{I^S(N^R - I^R)}{N^R} \left(1 - e^{-N_e^R f_{RB}}\right) \end{array} \right. \quad (3.3)$$

$$\left\{ \begin{array}{l} I^S(0) = 1, I^R(0) = 0. \end{array} \right. \quad (3.4)$$

Initially, one S plant is infected, but none of the R plants is. S and R plants can be infected

either by S or R plants, at a basic contact rate β_0 , as described previously. The mechanism of survival of viruses through the bottlenecks is modelled similarly as in eq. 3.1 for infection of S plants (last term in eq. 3.2) and for infection of R plants when the vector comes from an R plant (first term of eq. 3.3). Indeed, as we assumed that the RB variant is present at a frequency of 1 in R plants (gene-for-gene interaction system, Flor (1971)), the vector transmitting viruses from an infected R plant to a healthy R plant is necessarily inoculating the RB variant to the targetted R plant, and we assume that the survival of at least one RB particle is sufficient to infect the R plant. The 'genetic drift' term for the infection of a R plant from a S plant is slightly different (term $1 - e^{-N_e^R f_{RB}}$ in eq. 3.3). In this case, the frequency of the RB variant in the source S plant is f_{RB} . We assume that the target R plant will be infected if and only if at least one RB variant survives the bottlenecks. We model this condition by the probability density of a Poisson distribution of mean $N_e^R f_{RB}$. Indeed, if in average N_e^R virus particles are transferred from one infected R plant to a healthy R plant, only $N_e^R f_{RB}$ of these virus particles correspond to the RB variant. The probability density that at least one RB particle is transmitted and survives all bottlenecks is found by taking the opposite of the event leading to zero RB variant surviving, given by $1 - e^{-N_e^R f_{RB}}$. More details about the modeling of the genetic drift process and the associated distributions will be given when describing the stochastic form of the model.

3.2.3 Stochastic model description

Continuous time Markov chains and birth processes were chosen for the stochastic form of the model (Brauer et al., 2008). As for the deterministic form, we follow the dynamics of the number of S and R infected plants along the cropping season, and only 'birth' of new infected plants is possible (no 'death' by recovery or uprooting). The transition rates π for the 'birth' of new infected plants, because of a contact between an infected plant and a healthy one, correspond to those defined for the deterministic form of the model. We define π for each possible contact, as follows.

- Contact from an infected S plant to a healthy S plant : $\pi^{SS} = \beta_0 \frac{I^S(N^S - I^S)}{N^p}$
- Contact from an infected R plant to a healthy S plant : $\pi^{RS} = \beta_0 \frac{I^R(N^S - I^S)}{N^p}$
- Contact from an infected R plant to a healthy R plant : $\pi^{RR} = \beta_0 \frac{I^R(N^R - I^R)}{N^p}$
- Contact from an infected S plant to a healthy R plant : $\pi^{SR} = \beta_0 \frac{I^S(N^R - I^R)}{N^p}$

β_0 represents a basic contact rate per plant due to the activity of the vectors feeding from one plant to the next. Time was modelled thanks to the Gillespie algorithm (Gillespie, 1977),

defining here the waiting time until the next infection attempt (Wilkinson, 2006). This waiting time is assumed to follow an Exponential distribution, as :

$$\Delta t = \text{Exp}(\pi^{SS} + \pi^{RS} + \pi^{RR} + \pi^{SR}).$$

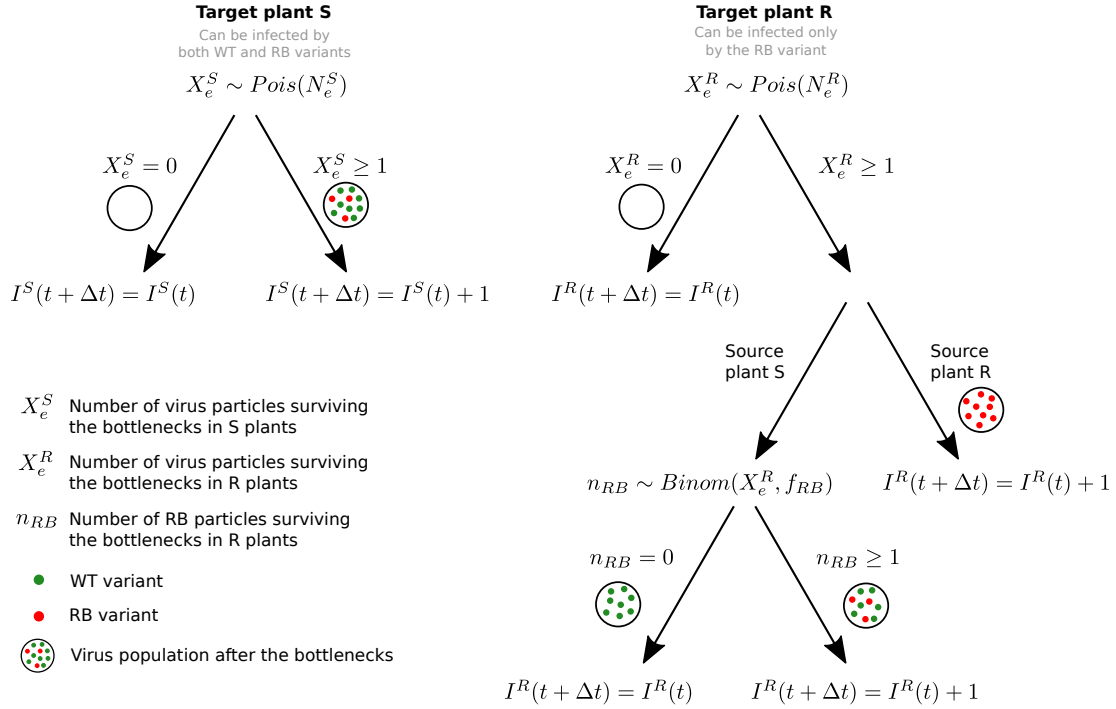


FIGURE 3.1 – **Implementation of the genetic drift process in the stochastic form of the model.** Two trees illustrate the model procedure for genetic drift implementation, when the target plant is susceptible (S, left tree) and when it is resistant (R, right tree). WT : wild-type; RB : resistance-breaking.

Once the waiting time is known, the varieties of the actual source and target plants are determined according to the probability of each event occurring $P_{ij} = \frac{\pi^{ij}}{\pi^{SS} + \pi^{RS} + \pi^{RR} + \pi^{SR}}$, with $i, j = S$ or R . As in classical Gillespie algorithm, this is achieved by a draw in a Uniform distribution, in looking at where it falls in the vector of cumulative probabilities, defined as $(0, P_{SS}, P_{SS} + P_{RS}, P_{SS} + P_{RS} + P_{RR}, 1)$. Then, the number of virus particles surviving the bottlenecks is drawn from a Poisson distribution, as $X_e^S \sim \text{Pois}(N_e^S)$ if the target plant is S, or $X_e^R \sim \text{Pois}(N_e^R)$ if the target plant is R (Fig. 3.1). If the target plant is S and if at least one virus particle survives the bottlenecks ($X_e^S \geq 1$), then the plant gets infected and $I^S(t + \Delta t) = I^S(t) + 1$. Otherwise (i.e. $X_e^S = 0$), the plant remains healthy and $I^S(t + \Delta t) = I^S(t)$. If the target plant is R, if no virus particle survives the bottlenecks ($X_e^R = 0$), then the plant remains healthy ($I^R(t + \Delta t) = I^R(t)$). If at least one virus particle survives ($X_e^R \geq 1$), two cases are to be distinguished. If the source plant is R, it means that the surviving virus particles are

necessarily of RB type, and hence the plant gets infected ($I^R(t + \Delta t) = I^R(t) + 1$). If the source plant is S, the plant will get infected if and only if at least one RB particle is part of the X_e^R surviving virus population. We assumed that the number of RB particles surviving the bottlenecks follows a Binomial distribution, as $n_{RB} \sim \text{Binom}(X_e^R, f_{RB})$. In simple terms, this draw in a Binomial distribution is equivalent to X_e^R draws in a Bernoulli distribution of mean f_{RB} , that determine, for each of the X_e^R surviving virus particles, whether it is of RB or WT type. This implies the assumption that virus particles are independent and do not interact with each other. From the mean fixed to f_{RB} , one can deduce our assumption that only genetic drift is at stake during the bottlenecks steps and that selection and mutation are neglected. The target R plant gets infected if and only if $n_{RB} \geq 1$. Otherwise it remains healthy. To highlight the match with the deterministic form of the model, let us note that a draw of X_e^R in a Poisson distribution of mean N_e^R followed by a draw in a Binomial distribution of parameters X_e^R and f_{RB} is indeed equivalent to directly drawing in a Poisson distribution of parameter $N_e^R f_{RB}$.

3.2.4 Model analysis

Measuring the yield increase in comparison with the fully S scenario

Our analysis of the benefit for farmers of using cultivars bearing quantitative resistance increasing the strength of genetic drift was based on the yield increase that it can provide. This quantity was measured thanks to the AUDPC, a good proxy of the yield losses caused by a pathogen (Fabre et al., 2012b; Jeger, 2004). Calculations were done for each :

- deployment strategy $\boldsymbol{\delta} = (\varphi, f_{RB}, N_e^R)$, involving a proportion φ of R plants, a value of f_{RB} characteristic of the choice of the R gene, and a virus effective population size N_e^R in R plants,
- and epidemiological context before deployment of R plants Ω_{int} , i.e. a value of the intensity of epidemics.

For one simulation, the AUDPC for a particular epidemiological context and deployment strategy is written $A(\Omega_{int}, \boldsymbol{\delta}) = \int_0^{n_d} [(1 - \varphi)p_i^S(t) + \varphi p_i^R(t)] dt$, summing out the weighted proportions of S and R infected plants, p_i^S and p_i^R , respectively. This quantity is compared with the one obtained with the reference field before deployment of R plants in the corresponding epidemiological context, $A_0(\Omega_{int}) = n_d \Omega_{int}$. For this purpose, we define the percentage of relative damage for a particular deployment strategy and epidemiological context as :

$$D(\Omega_{int}, \boldsymbol{\delta}) = 100 \times A(\Omega_{int}, \boldsymbol{\delta}) / A_0(\Omega_{int}). \quad (3.5)$$

For example, a value of relative damage D of 30% means that deploying R plants according to the strategy δ in the epidemiological context Ω_{int} reduces the number of infected plants to 30% compared to the crop damages before deployment of R plants (in the same epidemiological context).

Evaluating the benefit of using genetic drift to increase yield

From the relative damage values we can assess the additional yield benefit provided by pyramiding quantitative resistance reducing the virus effective population size with a qualitative resistance, compared to a R cultivar without such quantitative resistance, i.e. with a susceptible genetic background, resulting in a large effective population size. We will denote δ_{Rs} the strategies using a cultivar with a monogenic resistance, i.e. plants carrying a qualitative resistance R and a susceptible genetic background s , and δ_{Rr} the strategies using a cultivar with a polygenic resistance, i.e. plants pyramiding a qualitative resistance R with quantitative resistance in their genetic background r . The reference value of virus effective population size N_e^R in R plants for strategies using a monogenic resistance was set as 10^4 (Tab. 3.1). For strategies using a polygenic resistance, we allowed N_e^R to vary between 1 and 100, thereby reducing the effective population size by a factor 100 to 10^4 . We evaluated the added value of increasing genetic drift for yield benefit by comparing the strategies with monogenic and polygenic resistance differing only by the N_e^R parameter, i.e. with the same proportion of R plants φ , same characteristic of the R gene f_{RB} and same epidemiological context Ω_{int} . Following Fabre et al. (2015), we define the additional relative benefit of the polygenic resistance strategies as :

$$\Delta(\Omega_{int}, \delta_{Rr}) = \bar{D}(\Omega_{int}, \delta_{Rs}) - \bar{D}(\Omega_{int}, \delta_{Rr}), \quad (3.6)$$

with \bar{D} being the mean relative damage over the n_{iter} simulation iterations for one set of parameter values. An additional relative benefit Δ of 20 means that using polygenic resistance reduces the number of infected plants by 20 percentage points compared to the crop damages with monogenic resistance, in the same epidemiological context Ω_{int} and with the same proportion of R plants φ and characteristic of the R gene f_{RB} .

Simulation design

The simulation design made use of varying values of four parameters : the intensity of epidemics Ω_{int} , the proportion of R plants φ , the choice of the R gene f_{RB} and the virus effective population size in R plants N_e^R . All the other parameters were set to a fixed value. The values

of all parameters are given in table 3.1.

Based on all provided estimations of effective population sizes experienced by viruses just after host-to-host transmission and during within host progression (e.g. Gutiérrez et al., 2012b; Rousseau et al., Submitted; Zwart and Elena, 2015, see introduction section), we considered that a range from 1 to 100 for the effective population size in R plants pyramiding quantitative resistance was realistic. On the opposite, the value of the virus effective population size in the absence of quantitative resistance, i.e. in S plants or in monogenic resistant plants, was set to 10^4 , so that genetic drift is negligible. The intensity of epidemics Ω_{int} varied between 0.1 and 0.9, generating epidemics with on average 10% (resp. 90%) of the plants infected along the season. The proportion of R plants φ varied from 0.05 to 0.95 (plus the fully S field, with $\varphi = 0$, handled by the deterministic model), covering a large panel of deployment of R plants possibilities. The frequency of the RB variant in S plants f_{RB} , took values from 10^{-8} to 0.5. Values in the range $[10^{-8}, 10^{-4}]$ (resp. $[10^{-4}, 0.5]$) correspond to qualitative resistance genes typically requiring two (resp. one) mutation to be broken down, depending on the fitness costs associated with these mutations (Fabre et al., 2012b, 2015). This parameter was defined by Fabre et al. (2012b, 2015) in a context of a large virus population size, i.e. with negligible genetic drift.

The duration of the cropping season was fixed to 4 months ($n_d = 120$ days), as in (Fabre et al., 2012b, 2015). The number of plants in the field was set to 10^3 , being the upper limit for reasonable computational times and a lower limit for usual agricultural fields. Finally, we performed 500 simulation iterations for each set of parameter values, as this value was found to be a good trade-off between computational time and the precision of the mean and standard deviation values of the AUDPC over several batches of simulations.

3.3 Results

We begin with a description of model outputs. Then the effects of the four studied parameters Ω_{int} , ϕ , N_e^R and f_{RB} , will be investigated on the additional relative benefits $\Delta(\Omega_{int}, \delta_{Rr})$ and on the relative damage $D(\Omega_{int}, \delta)$.

3.3.1 Epidemic dynamics

When the reference intensity of epidemics in the fully S field $\Omega_{int} = 0.3$, the epidemic took off around 40 days after sowing and 95% of the plants were infected at the end of the season

(Fig. 3.2). Deploying the monogenic resistance cultivar ($N_e^R = 10^4$) at a proportion of $\varphi = 0.8$, with a resistance gene leading to a frequency of the RB variant in S plants of $f_{RB} = 0.01$, slightly reduced the epidemics in the S plants, leading to 86% of S plants infected at the end of the season, in average. In the R plants, the epidemic took off approximately at the same period, i.e. around 40 days after sowing, and the season ended up with 85% of R plants infected in average. Hence, the RB variant quickly invaded the field. Overall, in both S and R plants, adding R plants without quantitative resistance only slightly decreased the epidemic, with an epidemic intensity decreasing from 0.3 to 0.25 and the mean proportion of infected plants at the end of the cropping season being approximately 10 percentage points less than in the fully S field. Adding quantitative resistance reducing the virus effective population size to $N_e^R = 5$ in R plants (polygenic resistance), drastically slowed down the epidemics. Indeed, at the end of the cropping season the proportion of infected plants dropped to 11% in S plants, 7% in R plants, and 8% in the field (S and R plants), in average. With this strategy, the intensity of epidemics was reduced to 0.02. This example shows the benefit that can be provided by pyramiding quantitative resistance increasing the strength of genetic drift in R plants. In that case, the qualitative resistance reduces the damage by $(0.3 - 0.25)/0.3 = 16.7\%$, and the pyramiding of quantitative resistance accounts for an additional $\Delta = 100 \times \frac{0.25-0.02}{0.3} = 76.7$ percentage points decrease.

Demographic bottlenecks bring stochasticity to the dynamics of virus populations, and in return also to the epidemics. Hence, it is important to look at the variability present in the epidemics. Yet, a stochastic model intrinsically generates variability, which can be evaluated by looking at deployment strategies of monogenic resistance, i.e. with negligible genetic drift. In fact, figure 3.2 shows that the stochastic model itself already produces a large variability in the dynamics of the proportion of infected plants, with the epidemic dynamic curves being widely spread through the y-axis representing the proportion of infected plants. The density of curves at the last day of the cropping season shows that a large proportion of curves (21%) are concentrated around a specific value given by the peak of density at 96-98% of infected plants. The deployment strategy of polygenic resistance also shows quite widely spread epidemic curves, but not as much as for the monogenic resistance. The reason for this is the strong decrease of the epidemics spread, leading to flat epidemic curves at around 0% of infected plants and hence generating a so-called saturation effect. Indeed, the peak of density at the last day of the season is located at around 0-2% of infected plants for 61.6% of the simulations.

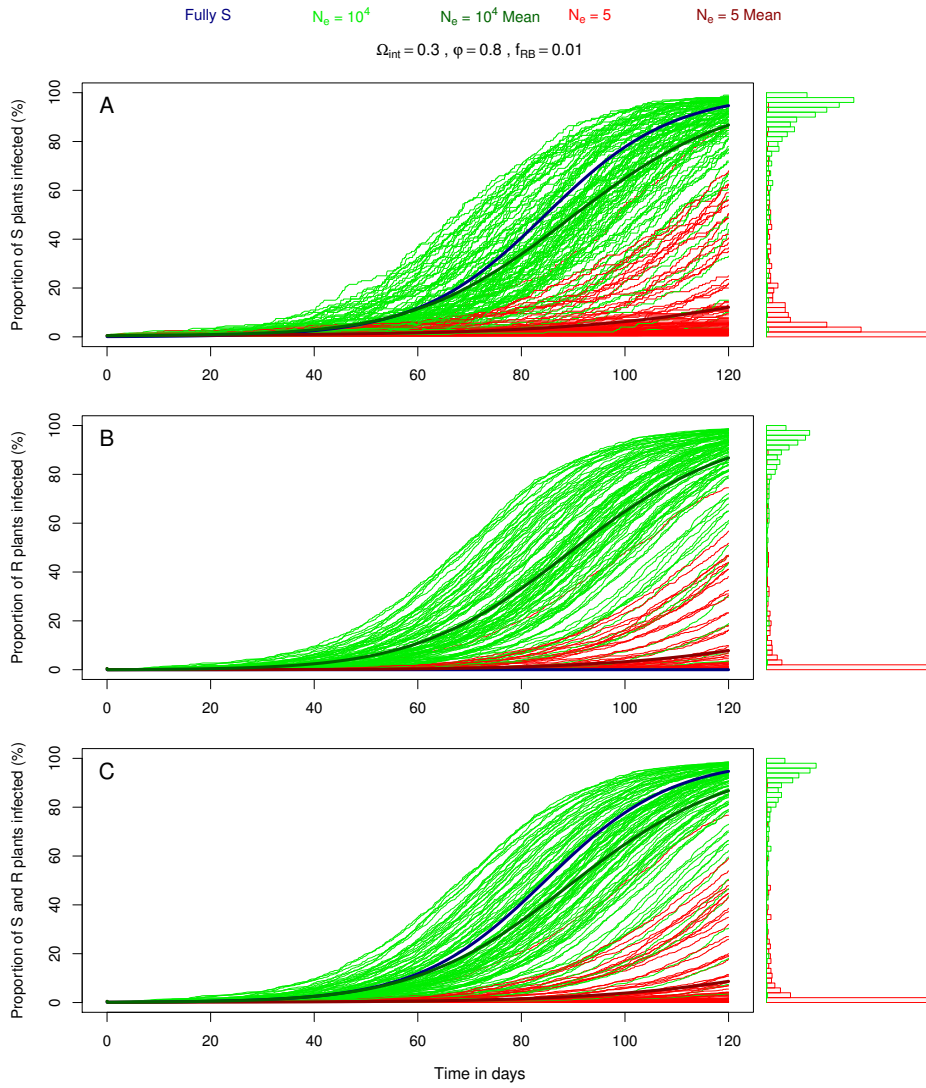


FIGURE 3.2 – **Dynamics of the epidemics.** Evolution of the proportion of infected plants during the cropping season, for susceptible S plants (panel A), for resistant R plants (B) and for both S and R plants (C). The blue thick curves are the outputs in a fully S field, obtained with the deterministic model, with a reference intensity of epidemics $\Omega_{int} = 0.3$. Green and red curves were obtained with the stochastic model, with a proportion of R plants $\varphi = 0.8$, a frequency of RB variant in S plants $f_{RB} = 0.01$ (500 simulations for each of the two sets of parameters, only 100 shown for clarity). The green thin curves are the outputs in a field with the monogenic resistance cultivar ($N_e^R = 10^4$). The green thick curves are the corresponding mean curves, with a time step of 0.1. The red thin curves are the outputs in a field with the polygenic resistance cultivar ($N_e^R = 5$). The red thick curves are similarly the corresponding mean curves. On the right of each graphic two histograms represent the proportion of infected plants at the last day of the cropping season with monogenic and polygenic resistance cultivars (green and red histograms, respectively).

3.3.2 Impact of the choice of the resistance gene

Impact on the additional relative benefit

We looked at the additional relative benefit Δ provided by adopting a polygenic resistance strategy δ_{Rr} compared to a monogenic resistance strategy δ_{Rs} as a function of the choice of the qualitative resistance gene f_{RB} , for all the values of the other parameters, Ω_{int} , φ and N_e^R

(Table 3.1). Analyses of the simulations showed that f_{RB} had a strong impact on the value of Δ that can be reached (Fig. 3.3). Only intermediate values of f_{RB} led to significant additional relative benefits, the extreme lower ($[10^{-8}, 10^{-6}]$) and higher (0.5) values remaining under 10 percentage points for the mean, and between -1 and 30 percentage points for the first and ninth deciles. On the contrary, when f_{RB} was between 10^{-5} and 0.1, the additional relative benefits remained above 17 percentage points for the mean, and between 0 and 85 for the first and ninth deciles. In other terms, pyramiding quantitative resistance only allows a substantial gain in epidemic control for genes characterized by intermediate f_{RB} values. The overall highest additional relative benefits are reached for $f_{RB} = 10^{-3}$, with a mean value of 54 percentage points. More generally, by looking at all the values of f_{RB} we can see that the first decile of Δ is either close to 0 percentage points, either above 0 percentage points. This means that for 90% of the set of parameters the additional relative benefits are either null or positive, but not negative. Hence, for the vast majority of cases, pyramiding quantitative resistance is either beneficial or neutral to increase crop yield, but it is not negative.

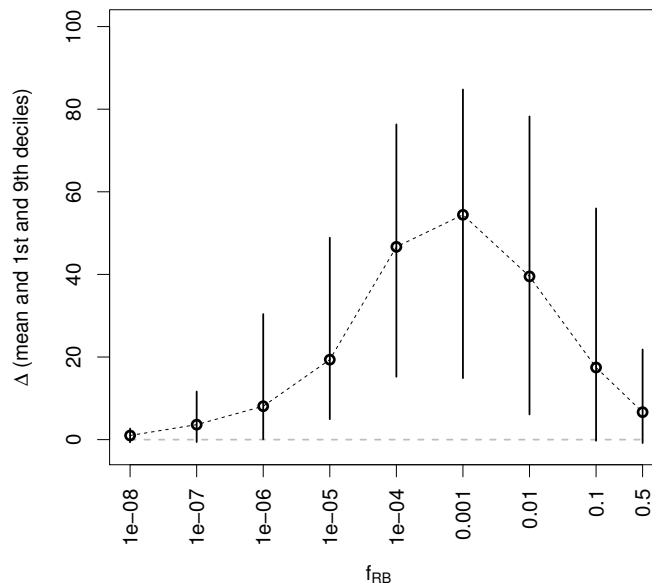


FIGURE 3.3 – **Additional relative benefit Δ as a function of the frequency of the RB variant in susceptible plants f_{RB} .** Mean (dots) and first and ninth deciles (segments) of Δ as a function of f_{RB} , all values of the other parameters - the intensity of epidemics Ω_{int} , the proportion of resistant (R) plants φ and the virus effective population size in R plants N_e^R - being combined. For illustration purposes, a black dashed line passing through all mean dots is represented. Also, a grey dashed line indicates the limit between positive and negative additional relative benefit.

Impact on the relative damage

Considering the relative damage D as a function of f_{RB} brings complementary results (Fig. 3.4). We represented the D variable for two fixed values of φ (0.2 and 0.8) and of Ω_{int} (0.2

and 0.8), and four values N_e^R , one corresponding to monogenic resistance strategies ($N_e^R = 10^4$), and the other three to polygenic resistance strategies ($N_e^R = 1, 10$ and 100). Figure 3.4 shows that stronger epidemic intensities generate larger relative damage in average. When φ is small (0.2), relative damages start at around 36-37% for the lowest value of f_{RB} and for low Ω_{int} (0.2), against 73-74% for high Ω_{int} (0.8). The pattern remains the same for all values of f_{RB} .

When φ is large (0.8), epidemics can get extinct (i.e. nearly all plants remain healthy over the cropping season), as the relative damage approaches 0% (epidemic-extinction case). The relative damage curves take off from the epidemic-extinction case for smaller values of f_{RB} when the intensity of epidemics is stronger. Typically the departure from this epidemic-extinction case is located at around $f_{RB} = 10^{-6}$ (resp. 10^{-4} for polygenic resistance strategies) when $\Omega_{int} = 0.2$ against $f_{RB} = 10^{-8}$ (resp. 10^{-6} for polygenic resistance strategies) when $\Omega_{int} = 0.8$.

A striking result is the effect of Ω_{int} on the variability of relative damages. The stronger the intensity of epidemics, the smaller the variability in the relative damages. Also, for the lowest value of Ω_{int} , relative damages can reach values larger than 100%. It means that the corresponding simulations generated more crop damage than in the fully S field. Yet, the mean relative damage does never get above 100%, showing that in average deploying R plants is beneficial.

The additional relative benefit is evaluated by the distance between the curves corresponding to the polygenic resistance strategies and the one of the monogenic resistance strategy. When f_{RB} is small, all the curves are close to each other at a small damage level, which means that no additional relative benefit is provided by adding quantitative resistance. This is especially true when Ω_{int} is low (0.2), as the curves stay close for larger values of f_{RB} . The result of low additional relative benefit when $f_{RB} = 0.5$ in figure 3.3 with higher damage level is here split in two cases. It is true when N_e^R is equal to 10 or 100, but less so when $N_e^R = 1$. In that latter case, it is true only when Ω_{int} is strong (0.8) and φ is small (0.2); otherwise relative damage still gets reduced. For larger N_e^R and intermediate f_{RB} values, the damage reduction is maximal when Ω_{int} is low (0.2) and φ is large (0.8). As in figure 3.3, we can see that intermediate values of f_{RB} lead to the largest reduction of relative damage when adding quantitative resistance. We can go further here, by noticing that a larger φ (0.8) leads to more additional relative benefit.

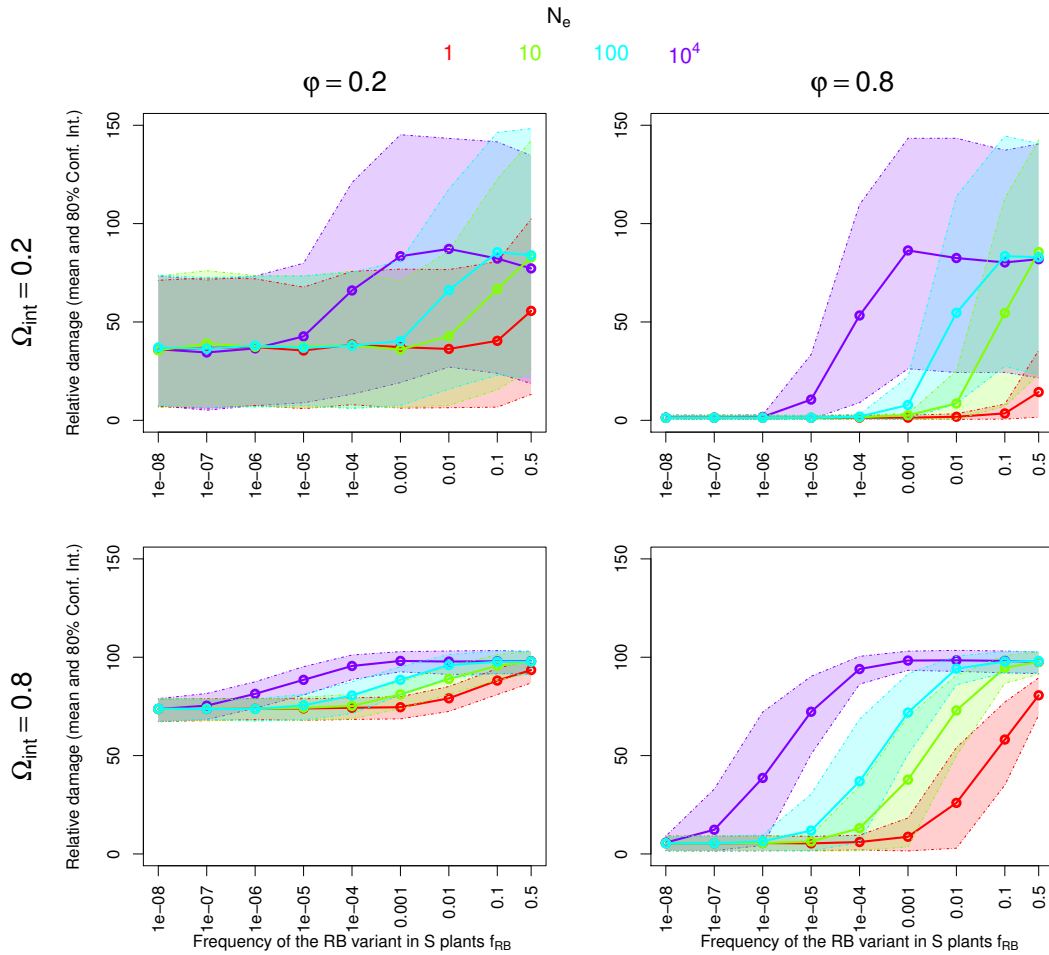


FIGURE 3.4 – **Relative damage as a function of the frequency of the RB variant in susceptible plants f_{RB} .** Mean (points and dotted lines) and 80% confidence interval (dashed lines and semi-transparent areas) on relative damage as a function of the frequency of the RB variant in susceptible plants (f_{RB}). Various viral effective population sizes are represented by different colors (one large N_e^R value (10^4) corresponding to monogenic resistance strategies, and three smaller values corresponding to polygenic resistance strategies). Each panel line corresponds to one value of the intensity of epidemic, and each panel column to one value of the proportion of resistant plants.

3.3.3 Global impact of all the factors

To finish with, we considered the combined effects of the four factors, Ω_{int} , f_{RB} , ϕ and N_e^R , on the additional relative benefits Δ (Fig. 3.5). For clarity, figure 3.5 presents only a subset of the parameter values : $\Omega_{int} = \{0.2, 0.5, 0.8\}$ and $f_{RB} = \{10^{-4}, 10^{-2}, 0.5\}$. When f_{RB} is small (10^{-4}) and Ω_{int} is small or intermediate (0.2 or 0.5), no visible effect of N_e^R can be noticed. Only ϕ had some effect, leading to higher additional relative benefits as ϕ gets larger. Increasing ϕ can provide up to 50-60 percentage points of additional yield benefit when $\Omega_{int} = 0.2$, and up to 80 percentage points when $\Omega_{int} = 0.5$. When Ω_{int} is strong (0.8), a slight effect of N_e^R appears, leading to higher additional relative benefit when N_e^R is small, for large ϕ values. Again, increasing Ω_{int} allows to reach higher levels of additional relative benefit, as in that

case the maximum value is in the range of 90-100 percentage points. When f_{RB} is intermediate (0.01), N_e^R does have an influence for any epidemic intensity, even generating J-shaped contour lines when Ω_{int} is small (0.2) or intermediate (0.5). The most efficient strategies then lie in the region where φ is large and N_e^R is small. When f_{RB} is large (0.5), the effect of φ blurs, and the areas corresponding to significant additional relative benefit (e.g. ≥ 10 percentage points) are reduced to small values of N_e^R (< 5). These areas tend to disappear as Ω_{int} increases, moving towards very small N_e^R (< 2) and large φ (> 0.6) values. The additional relative benefit can reach up to 70 percentage points when $\Omega_{int} = 0.2$ and this maximum decreases to 20 percentage points when $\Omega_{int} = 0.8$.

Overall, Δ hardly ever went below 0 percentage points in figure 3.5. The minimal Δ value was of -4.4 percentage points, but 75% of negative Δ were greater than -1 percentage points.

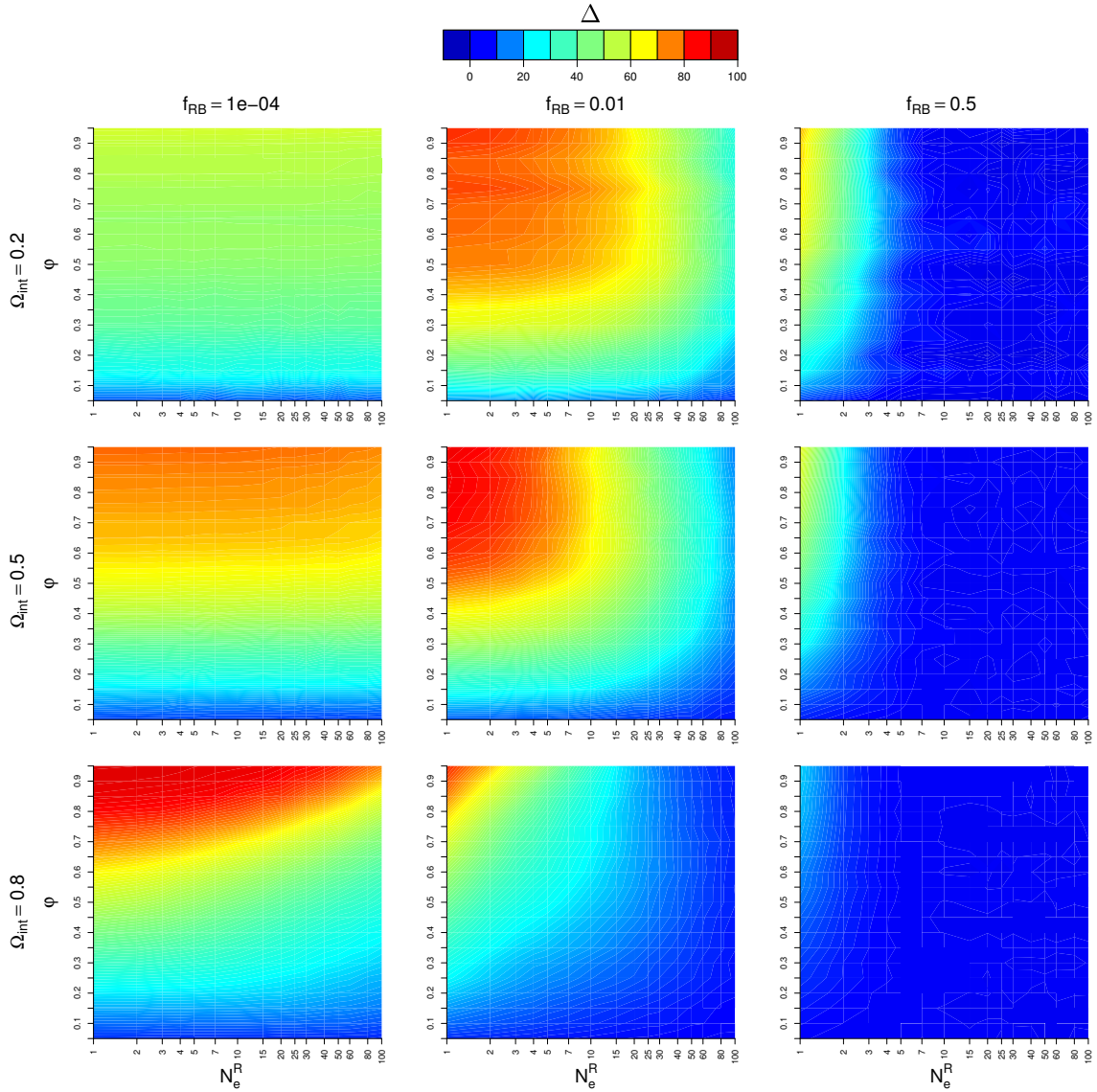


FIGURE 3.5 – Additional relative benefit Δ as a function of four parameters representative of the host, the virus, the vectors and the field. Contour plots representing Δ as a function of the proportion of resistant plants φ on the y-axis and of the virus effective population size N_e^R in R plants on the x-axis. Panel lines represent three contrasted values of the intensity of epidemic Ω_{int} (0.2, 0.5 and 0.8). Panel columns represent three contrasted values of frequencies of the RB variant in susceptible plants f_{RB} (10^{-4} , 10^{-2} and 0.5). The color legend for Δ is indicated on top of the graphics.

3.4 Discussion

The epidemiological model presented in this paper is to our knowledge the first proposed to analyze the impact of genetic drift on plant disease resistance durability. It is rather simple but relevant, and it provides a number of results from which we could identify the effects of genetic drift and from which basic guidelines for farmers can be drawn.

3.4.1 Global impact of the choice of the resistance gene on the additional relative benefit

The first effect put into light concerned the global impact of the choice of the R gene, characterized by the frequency of the RB variant in S plants (f_{RB} parameter), on the additional relative benefit. More precisely, the additional relative benefit followed a bell-shaped curve as a function of f_{RB} , with both ends corresponding to very low additional relative benefit. Only intermediate f_{RB} values provided substantial additional benefit.

This occurs firstly because when the RB variant frequency in S plants is small ($f_{RB} \leq 10^{-6}$), the resistance is hardly breakable, even without quantitative resistance in the genetic background of R plants ($N_e^R = 10^4$). Such resistance genes typically require numerous mutations for the virus to break them and high associated fitness costs in S plants (this is why f_{RB} is very low) (Carrasco et al., 2007; Fabre et al., 2012b; Fraile et al., 2011; Sanjuán, 2010). In those cases, the probability that at least one RB virus particle survives the bottlenecks in R plants is at most equal to $1 - e^{-10^4 \times 10^{-6}} = 10^{-3}$ (when $f_{RB} = 10^{-6}$). It means that in average one R plant gets infected after 10^3 trials when the vector comes from an infected S plant. In other words, such a qualitative resistance already leads to satisfactory results in terms of yield benefit without pyramiding it with quantitative resistance. Increasing the strength of genetic drift would not bring significantly more efficacy. Such cases of highly durable qualitative resistances have been reported in agroecosystems, as for the *Tm2* gene in tomato and *N*-gene in tobacco against TMV (Parlevliet, 2002), or for the *pvr2²* resistance allele in pepper against PVY (Ayme et al., 2007; Fabre et al., 2012a).

Second, when the RB variant frequency in S plants is large ($f_{RB} = 0.5$), the situation is as if the resistance was already broken down because the RB variant is initially present at a high frequency in the field. Indeed, such a value of f_{RB} typically corresponds to a situation where bearing the RB allele does not induce any fitness cost in S plants and a unique mutation is sufficient for the virus to become RB (Fabre et al., 2012b). In that case, even the narrowest bottlenecks in R plants ($N_e^R = 1$) lead to a probability that at least one RB virus particle survives equal to 0.39. This means that in average after 2-3 trials a R plant gets infected when the vector comes from an infected S plant. It should not take long before the first R plant gets infected, after what infection of R plants becomes easier as the route of infection "from an infected R plant to a healthy R plant" is then opened and yields higher infection success probabilities. In other words, and more generally, the infection of the first R plant is a necessary step which is the most difficult to achieve, but once it is, the RB variant can invade the other R

plants easily by transmission from infected R plants to healthy R plants (Brown, 2015). Cases of poorly durable resistances have also been reported in agroecosystems, as for the *Tm1* gene in tomato against TMV (Parlevliet, 2002), or the *Tsw* gene in pepper against the *Tomato spotted wilt virus* (TSWV, genus *Tospovirus*) (Moury et al., 1997).

To finish with, the global impact of f_{RB} , intermediate values of this parameter can provide the largest additional benefits because the qualitative resistance is neither unbreakable nor easily broken down. Hence, in those common intermediate cases (Carrasco et al., 2007; Fabre et al., 2012b), quantitative resistance controlling bottleneck sizes can strengthen a qualitative resistance by decreasing the probability of success of infection attempts from S plants to healthy R plants. More generally, in agroecosystems all cases of durability of qualitative resistances against viruses have been reported, varying between not durable, intermediate and highly durable (García-Arenal and McDonald, 2003).

3.4.2 Impact of the four factors

The model shows an interesting combined effect of the intensity of epidemics Ω_{int} and the proportion of R plants φ on relative damage D . When the majority of plants is S ($\varphi \leq 0.5$), stronger epidemic intensities generate larger relative damage in average (Fig. 3.4). Indeed, strong intensities of epidemics are associated with high contact rates β_0 . For example, moving from $\Omega_{int} = 0.2$ to $\Omega_{int} = 0.8$ in a fully S field increases the basic contact rate β_0 from 0.095 to 0.288. This means that the waiting time for a new contact event from an infected plant to a healthy plant will be shorter, hence more trials of infection will be attempted during the cropping season. This leads to higher successful infection probabilities along the season, and hence to larger relative damages, even if genetic drift is strong.

When the majority of plants is R ($\varphi \geq 0.5$), this phenomenon is not visible anymore, but another one comes into play. Higher epidemic intensities lead the additional relative benefit to take significantly positive values for lower RB variant frequencies f_{RB} in S plants (Fig. 3.4). Two conclusions can be drawn from this result. First, the fact that the relative damage is located at around 0% for all N_e^R values and for the lowest f_{RB} values teaches us that the epidemic was already extinct with the monogenic resistance alone. Hence, in those cases, the polygenic resistance could not provide anything more in terms of yield. Second, the fact that more additional relative benefit is achieved for lower f_{RB} values when epidemic intensity is higher is directly linked to the following points. When the epidemic intensity is high, the relative damage of the monogenic resistance strategy takes off from the epidemic-extinction

cases for lower f_{RB} values (Fig. 3.4). This follows from the fact that more trials of infection are attempted when the epidemic intensity is higher, leading to higher overall successful infection probabilities. In return, as the epidemic does not get extinct anymore, genetic drift can provide benefit again.

The variability of the relative damages decreased with the intensity of epidemics. In fact, the stronger the intensity of epidemics, the sooner all plants get infected, a point beyond which the dynamics necessarily remains constant. This explains why there was less variability in the epidemic dynamics, and in return in the relative damage.

We reported cases where the relative damage got above 100% when the intensity of epidemics is small. This means that the plant damage was larger when R plants were deployed in comparison with deterministic epidemics in a fully S field. It does not necessarily imply that deploying R plants was harmful, as one major change from the fully S field simulations to the resistance deployment simulations was the use of a stochastic model, which can lead to such a result simply because of the variability in epidemic dynamics. Anyway, the mean relative damage never went above 100% for the set of parameters used in figure 3.3, and the additional relative benefit hardly ever went below 0 percentage points for the set of parameters used in figure 3.5, showing that in average deploying R plants is beneficial. Hence we can confidently put forth the hypothesis that cases of negative additional relative benefit were due to the intrinsic stochasticity of the model and not to genetic drift.

Several outlines emerge from figure 3.5. When the RB variant frequency in S plants is small, no significant effect of bottlenecks size (N_e^R) is noted, only an effect of the proportion φ of R plants. In such a case, the best strategy to increase the yield benefit is to deploy the largest proportion of R plants. Conversely, when the RB variant frequency in S plants is large, no significant effect of the proportion of R plants is noted, but an effect of bottlenecks size. The best strategy to adopt is then to choose an R cultivar with quantitative resistance leading to the smallest virus effective population sizes. When the RB variant frequency in S plants is intermediate, both the proportion of R plants and bottleneck sizes are important, and the best strategy corresponds to a large proportion of R plants and a small bottleneck size. However, when the intensity of epidemics is strong, the effect of both parameters becomes less visible for all R genes characteristics. When f_{RB} is small and Ω_{int} is large, it is also preferable to choose quantitative resistance leading to small bottlenecks sizes. For example, when the proportion of R plants is 70%, reducing virus bottleneck sizes can provide up to 40 percentage points of additional relative benefit. For strong epidemic intensities, the optimal areas are

drastically reduced as the frequency of the RB variant in S plants f_{RB} increases, a feature well explained by the increasing number of infection attempts combined with increasing probabilities of transmitting the RB variant from an infected S plant.

Overall, the model always predicts higher additional benefit provided by pyramiding quantitative resistance when the targeted bottlenecks are narrowest, when the R cultivar is deployed at large proportions and when the qualitative resistance gene corresponds to intermediate f_{RB} . Furthermore, more additional benefits are achieved when the epidemic intensity is lower. McDonald and Linde (2002a) and García-Arenal and McDonald (2003) included the pathogen effective population size as a risk factor for the evolutionary potential of pathogens confronted with a qualitative resistance in plants. As for our conclusions, these authors and Zhan et al. (2015) advise to adopt strategies minimizing the pathogen effective population size to reduce the risk of resistance breakdown. Regarding f_{RB} , which decreases with the number of mutations required for resistance breakdown, the conclusion we draw is not straightforward. Previous studies reported an increase of resistance durability with the number of mutations needed for resistance breakdown (Fabre et al., 2009; Harrison, 2002; Lecoq et al., 2004), and with the fitness cost associated with these mutations (Fraile et al., 2011; Janzac et al., 2009, 2010; Leach et al., 2001). Here, if f_{RB} is low, the resistance gene is highly efficient and genetic drift is useless. The utility of genetic drift is revealed at intermediate f_{RB} values. Intermediate proportions of R plants φ were predicted to be optimal most of the time for yield benefit in several studies (Fabre et al., 2012b; Ohtsuki and Sasaki, 2006). Here, we predict optimal additional benefit at large φ , but we will see in the next paragraph that this may be due to a limitation of the model framework.

3.4.3 Model-imposed link between the proportion of resistant plants and the intensity of genetic drift

The modeling framework developed imposes a link between the proportion of R plants and the intensity of genetic drift because the only option we tested was to introduce quantitative resistance controlling the bottleneck sizes imposed to virus populations in the genetic background of the R cultivar. The bottleneck size at the individual plant scale, N_e^R , is important, but it is the product between N_e^R and the proportion of plants carrying the corresponding quantitative resistance that determines the intensity of genetic drift at the field scale. This is probably why we did not observe significant additional relative benefit when the proportion of R plants was small, while the highest additional relative benefits corresponded to the largest proportion of

R plants. Also, simulating over one cropping season only leaves less time for the pathogen to break the qualitative resistance down. The breakdown must be very fast so that deploying a large proportion of R plants would not be beneficial. Hence, the model predicts larger φ values to lead to higher resistance durability. Yet, the link could be broken, for example by allowing quantitative resistance to be introduced in both the R and the S cultivars. This could require the association of four plant varieties in the field : the S cultivar either with a susceptible genetic background or quantitative resistance, and the R cultivar either with a susceptible genetic background or with quantitative resistance. The intensity of genetic drift at the field scale would then not be dependent on the proportion of R plants, and we could see if the positive link between this latter parameter and the additional benefit is kept or if another shape of link emerges.

Using a reduced set of parameters, we assessed the additional relative benefit provided by introducing the quantitative resistance in the S cultivar and using only the monogenic resistance for the R cultivar (data not shown). We compared this alternative strategy with the current one presented in this article, i.e. using the S cultivar without quantitative resistance and the R cultivar with quantitative resistance, i.e. polygenic resistance. In all, the alternative strategy led to higher additional relative benefit than with the current strategy in 23% of the set of parameters tested, but for 75% of those cases the additional benefit was less than 5 percentage points superior to the additional benefit provided by the current strategy using polygenic resistance. Cases where introducing the quantitative resistance in the S cultivar were better than the current strategy of polygenic resistance were usually associated with a very low RB variant frequency in S plants, i.e. typically for resistance genes hardly breakable, and with a small proportion of R plants in the field. The effect of f_{RB} is easily understandable. When the resistance gene is hardly breakable (low f_{RB} values), the qualitative resistance is already very efficient and durable by itself, so that the R plants are at low risk of infection. Hence, introducing quantitative resistance in S plants seems a good strategy to slow down the infection spread through the S plants. The link with the proportion of R plants is observed again, but obviously with a correlation opposed to the one for the strategy using polygenic resistance.

3.4.4 Future developments of the model

Implementation of more complex evolutionary pathways for resistance breakdown

In this study we focused on the epidemiological dimension rather than on the virus evolution. This latter aspect can nevertheless be important, and it would be of great interest to include it in future research. Notably, Quenouille et al. (2013) showed that, for polygenic resistances, the mutational pathways for the WT variant to become RB were fewer and more complex, generally requiring more mutations than for the monogenic resistances. Additionally, the virus may also adapt to the quantitative resistance. Recent papers showed both aspects. Several studies observed the adaptation of pathogens to quantitative resistance, as for the PVY to a quantitative resistance reducing virus accumulation in pepper (Montarry et al., 2012), for *Venturia inaequalis* to quantitative resistance in apple (Caffier et al., 2014), or for *Plasmopara viticola* to quantitative resistance in grapevine (Delmotte et al., 2014). Other studies observed no significant adaptation to quantitative resistance, as for *Leptosphaeria maculans* to quantitative resistance pyramided with a qualitative resistance (*Rlm6* resistance gene) in *Brassica napus* cultivars (Brun et al., 2010).

Introducing trade-offs between benefits and drawbacks of population bottlenecks

Deleterious effects of narrow bottlenecks have been reported in nature, as for *Trypanosoma cruzi*, the agent of Chagas disease (Levy et al., 2015). The authors argued that the seasonal drastic reduction in guinea pig hosts populations in Peru led to a high prevalence of *T. cruzi* in the insect vector population (triatomine), because vectors feeding on a restricted number of hosts have a higher probability of feeding on infected ones, and in return, this increased the risk of Chagas disease in households. Additionally, Miyashita and Kishino (2010) conducted an experimental study on plant viruses in which several mutational events were required for resistance breakdown. They argued that narrow bottlenecks during cell-to-cell movement in plants helped isolating variants with adaptive mutations from not adapted ones, allowing selection to operate more efficiently. Zwart and Elena (2015) strengthened that point in the case where the fitness of the transitional mutants is lower than the fitness of the wild-type and the adapted variants (reciprocal-sign epistasis effect between mutations). The successive bottlenecks during the virus cell-to-cell colonization can allow the fixation of the transitional mutants, enabling the virus to acquire the remaining mutations necessary for resistance breakdown.

Our modeling framework was built in a way such that genetic drift could not be harmful, because the probability of successful infection of a R plant is a strictly increasing function of

N_e^R . One way to improve the model on that point would be to rely on the notion of time to fixation of the RB adapted variant from the population genetics theory, in addition to the notion of probability of fixation of the RB variant that we used. When the effective population size is large, the time to fixation is long but the probability of fixation is strong, whereas when the effective population size is small, the time to fixation is short but the probability of fixation is weak (Gerrish and Lenski, 1998; Sniegowski and Gerrish, 2010). Introducing a waiting time, during which the plant is not infected, until the fixation of the RB variant in R plants when the effective population size is large (typically $N_e^R = 100$), would be a good way to slow down epidemics when genetic drift is not strong. This trade-off between the probability of fixation and the time to fixation could counteract the positive relationship between N_e^R and resistance durability.

3.4.5 Extending the spatial and temporal scales of the model

The model was developed and analyzed for one field considered during one cropping season. It has often been argued that the best scale of deployment of a control strategy is the scale at which epidemics naturally occur (Dybiec et al., 2004; Fabre et al., 2012b; Gilligan, 2008). This scale would correspond to a landscape composed of several fields for crop epidemics. In this contribution, we focused on the field scale because of heavy computational times. Nevertheless, drawing conclusions at the field scale constitutes an appropriate first step, without the complex additional characteristics corresponding to a landscape scale approach (several fields, virus reservoir in wild-type plants, etc.).

Another limitation of our model is the short time period over which we simulated the epidemics : one cropping season. The durability of a resistance is a criterion best measured over a long period, i.e. over several years (Fabre et al., 2012b, 2015; Johnson, 1984). We chose not to simulate over several years because we wanted to keep our model simple. Nevertheless, this aspect should be included in future research. More generally, even if we simulated epidemics over one cropping season only, the yield improvement provided by genetic drift is already a proxy of the associated potential increase in durability.

Overall, such long time period and large scale simulations have already been achieved to answer other questions linked to resistance durability, notably by Fabre et al. (2012b, 2015). We could use those models as a basis to extend our model on those points.

3.4.6 Concluding remarks

Our study is a first step towards the demonstration that integrating population genetics principles to minimize the evolutionary potential of plant pathogens, by playing on their effective population size, can guide disease resistance management strategies towards resistance durability and epidemic control (McDonald and Linde, 2002a; Zhan et al., 2015). The simple model developed and analyzed here already provides insightful guidelines for breeders, by showing the benefit of quantitative resistance reducing virus bottleneck size, particularly when the combined qualitative resistance is not highly durable and the RB variant is not already present at a high frequency in the agroecosystem. This model also provides farmers with optimal deployment strategies. However, several limiting points of the model were indicated in this discussion, most importantly the small space and time scales of simulations, and should be accounted for in a future improved version.

Acknowledgements

The authors would like to thank the cluster MIGALE from the Bioinformatic platform of INRA Jouy en Josas for being the support of all simulation runs presented in this paper. The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under grant ANR-13-BSV7-0011 (project FunFit).

Chapitre 4

Joint Estimation of Effective Population Size and Selection Coefficient without Neutral Markers : Method Validation and Application to the Experimental Evolution of Viruses

Elsa Rousseau^{1¶*,2,3}, Benoît Moury³, Ludovic Mailleret^{1,2}, Rachid Senoussi⁴, Alain Palloix⁵,
Vincent Simon^{3,6}, Sophie Valière^{7,8}, Frédéric Grogard¹, Frédéric Fabre^{9¶*}

¹ Biocore Team, INRIA, Sophia Antipolis, France

² INRA, Univ. Nice Sophia Antipolis, CNRS, UMR 1355-7254 Institut Sophia Agrobiotech, Sophia Antipolis, France

³ UR PV, INRA, Montfavet, France

⁴ UR BioSp, INRA, Avignon, France

⁵ UR GAFL, INRA, Montfavet, France

⁶ UMR BFP, INRA, Villenave d'Ornon, France

⁷ GeT-PlaGe, INRA, Genotoul, Castanet-tolosan, France

⁸ UAR DEPT GA, INRA, Castanet-Tolosan, France

⁹ UMR SAVE, INRA, Villenave d'Ornon, France

¶These authors contributed equally to this work.

* Corresponding authors

elsa.rousseau@inria.fr (ER)

frederic.fabre@bordeaux.inra.fr (FF)

This article has been submitted to PLoS Genetics on March 9th 2016.

Abstract

The simultaneous estimation of effective population size N_e and selection coefficient s from time-sampled data is challenging if both genetic drift and selection are strong and no neutral markers available. This is a common situation for experimental evolution studies with microorganisms. We have developed a mechanistic-statistical model for estimating these parameters from genetic time series data generated in evolve-and-resequence experiments. Numerical simulations of the dynamics of competing genotypes in asexual haploid Wright-Fisher populations subjected to contrasting genetic drift ($N_e \in [20, 5000]$) and selection ($|s| \in [0, 0.1]$) regimes were used to test the model. This model was then applied to data for competition between five virus variants in 15 host plant genotypes. The frequencies of the variants were determined at six time-points, in eight plants per time point for each host environment.

The numerical experiments validated the method and made it possible to obtain bias-adjusted estimators of s for each competing genotype and of the dynamics of N_e over the time course of the experiment. The real-life experiment showed that the virus populations experienced either stochastic or deterministic evolution, depending on host genotype and that N_e and s were genetically controlled by the host genotype.

This method constitutes an advance in the joint estimation of key population genetic parameters from time-sampled data in conditions in which no neutral markers are available and both selection and genetic drift are strong. This is a situation frequently encountered in studies of microorganisms. The observation of very contrasted genetic drift between plant genotypes may have practical implications for slowing viral emergence through the careful choice of plant cultivars.

Key words

Genetic drift, Experimental evolution, High-throughput sequencing, Maximum-likelihood, MiSeq Illumina, Plant virus, Selection, Virus evolution

Author Summary

A growing number of experimental evolution studies used “evolve-and-resequence” approach in which the genomes of experimental populations are repeatedly sequenced over space and time. The resulting data are particularly appropriate to study basic evolutionary forces such as selection and genetic drift. However, disentangling the effects of selection and genetic drift still remains challenging when no neutral markers are available, a situation which is frequent for microbes such as viruses, due to their small, tightly packed genomes. Here, we propose and validate a method that does not require neutral markers to estimate jointly selection coefficients and effective population sizes from molecular data issued from experimental evolution. This method is widely applicable to evolve-and-resequence experiments. We applied our method to high-throughput sequencing data of virus populations infecting several plant genotypes. Analyses revealed that plant viruses experienced a considerable diversity of selection and genetic drift regimes in closely related host genotypes. Results also indicated that genetic drift and selection imposed by the plants to the virus population were heritable plant traits. These findings pave the way for the breeding of plant varieties exposing viruses to stronger genetic drift and/or smaller selection effects, thus, slowing down viral adaptation.

4.1 Introduction

Evolution in isolated populations results from interplay between several forces, including mutation, selection, and genetic drift (Charlesworth, 2009; Rouzine et al., 2001). Mutation creates genetic diversity within a population. Subsequent selection and genetic drift then drives the evolution of diversity within the population. Selection is a deterministic force that increases the frequency of the fittest variants at the expense of the weakest ones. It can be characterized by the selection coefficient s , commonly calculated, at a specific locus, as the relative difference in fitness conferred by two alleles. Genetic drift, unlike selection, acts equally on all variants. It is the outcome of random sampling effects between generations, resulting in stochastic fluctuations in variant frequencies (Charlesworth, 2009). The strength of genetic drift is frequently evaluated by calculating effective population size N_e (Charlesworth, 2009). The N_e concept was introduced by Sewall Wright in 1931. Effective population size is defined as the size of a panmictic population of constant size with non-overlapping generations that would display the same degree of randomness in allele frequencies as the population studied (Wright, 1931). N_e is often much lower than the census population size (Gutiérrez et al., 2012a; Vucetich et al., 1997), but it can be seen as its evolutionary analog (Waples et al., 2014). Indeed, ecological processes, such as competition and population growth, are governed by census population size, whereas evolutionary processes, such as genetic drift and selection, are primarily governed by effective population size (Waples et al., 2014). When N_e is small, sampling effects are magnified between generations, and allele frequencies therefore fluctuate strongly. For populations varying in size over time, the effective population size over a given number of generations can be approximated by the harmonic mean \bar{N}_e of effective population sizes at each generation. This approximation holds provided that the number of generations is much smaller than \bar{N}_e (Caballero, 1994; Kimura, 1970; Kimura and Crow, 1963; Lande and Barrowclough, 1987; Wright, 1939) and mutation can be neglected (Motro and Thomson, 1982). Population size may vary over time due to bottlenecks, which are common in natural populations. As they greatly decrease population size, they have a disproportionate effect on the overall value of \bar{N}_e (Charlesworth, 2009; Lande and Barrowclough, 1987). The product $N_e \times s$ estimates the effectiveness of the deterministic force of selection relative to the stochastic force of genetic drift (Charlesworth, 2009; Rouzine et al., 2001). If $N_e \times s \ll 1$, then genetic drift predominates over selection and evolution is mostly stochastic. If $N_e \times s \gg 1$, then selection becomes effective and evolution is mostly deterministic (Rouzine et al., 2001).

In recent decades, various approaches based on moment (Nei and Tajima, 1981; Waples,

1989) or likelihood (Anderson et al., 2000; Berthier et al., 2002; Williamson and Slatkin, 1999) methods have been proposed for estimating N_e . All require the genetic markers studied to be under neutral selection. Various methods have also been proposed for detecting selection and estimating selection coefficients. These methods require at least some prior information about N_e (e.g. Vitalis et al. (2014)) or assume that genetic drift is negligible (e.g. Illingworth et al. (2012)). However, in the absence of neutral markers and without prior estimates of N_e , both selection and genetic drift must be taken into account, as these two forces act simultaneously in evolution. This greatly complicates the estimation of N_e and s . Only a few methods have been proposed to date for the joint estimation of N_e and s from time-sampled data (see Foll et al. (2014b) and Malaspinas (2016) for a review). For large effective population sizes (typically $N_e > 5000$) and small selection coefficients (typically $|s| < 0.01$), several likelihood methods based on diffusion approximation of the Wright-Fisher model (Charlesworth, 2009; Foll et al., 2014b) are available (Bollback et al., 2008; Malaspinas et al., 2012; Mathieson and McVean, 2013; Steinrücken et al., 2014). In the situations in which these methods are valid, the ranges of N_e and s values obtained are rather restrictive for many microorganisms, particularly for those of our experimental model, viruses (Elena et al., 2011; Fabre et al., 2012a; French and Stenger, 2003; García-Arenal et al., 2003; Sacristán et al., 2003). Foll et al. (2014c) recently proposed the use of approximate Bayesian computation (ABC) for joint estimation of the N_e and s of a Wright-Fisher model. Their method can deal with both weak and strong selection ($|s| \simeq 0.1$) regimes, but it requires at least 90% of neutral genetic markers. This method is much better than those that preceded it, but the requirement for such a high proportion of neutral markers cannot always be satisfied, particularly for viruses, which have few neutral markers in their genomes (Elena et al., 2014; Sanjuán et al., 2004). Terhorst et al. (2015) recently used a Gaussian process to approximate a Wright-Fisher model with selection. Their method detects, locates and estimates the fitness of a selected allele (with $|s| \in [0.01, 0.1]$) from several linked markers. It has two requirements : (i) allele frequencies must not be too close to loss or fixation and (ii) the effective population size must be moderate ($N_e \simeq 1000$). The authors assumed that N_e was generally known, but they also tested the ability of their method to estimate N_e alone (*i.e.* assuming that selection coefficients are known).

We present here a mechanistic-statistical model based on deterministic Lotka-Volterra equations and stochastic Dirichlet-multinomial processes for the joint estimation of selection and genetic drift, whether these forces are strong or weak, in the absence of neutral markers. This method is particularly appropriate for experimental evolution studies. The parameters of

the model (i.e. selection coefficients for each allele and effective population size at each time-point) can be estimated from time-sampled data representing changes in allele frequencies in a population by maximum likelihood methods. Numerical simulations of a haploid Wright-Fisher model with mutation, selection and genetic drift (Zanini and Neher, 2012a) were used to validate this model. These numerical simulations typically mimic the time-series datasets obtained in evolve-and-resequence experiments (Turner et al., 2011) with high-throughput sequencing (HTS), a tool increasingly used in studies of microbial evolution (Brockhurst et al., 2011). The simulations covered an extensive range of N_e and s values, to make it possible to evaluate (i) the precision and bias of the estimates and (ii) the limitations of the model. We also describe a case study consisting of a real-life biological experiment on competing variants of *potato virus Y* (PVY) infecting various pepper genotypes. The same population of virus variants was used to inoculate various plant genotypes. A large range of evolutionary patterns was observed, from almost deterministic to almost stochastic evolution regimes. We estimated the selection coefficient of each variant in each genotype and changes in effective population size over time during the colonization of the plant by the virus.

4.2 Results

4.2.1 Overview of the method

We propose a method for the analysis of time-sampled data representing the frequency dynamics of a set of genotypes belonging to an asexual, haploid population, competing in a given environment. Genotypes are observed at different time-points in isolated populations derived from the same parental population. We will present the model below in the context of virus populations invading their host plants, although this model could be applied to a much broader range of situations. Virus variants are genetically closely related.

The mechanistic-statistical model we built to study these data combines deterministic and stochastic aspects. It was developed from the work of Fabre et al. (2012a) and will be named thereafter the Lotka-Volterra Dirichlet-multinomial (LVDM) model. The parameters and variables of the model are summarized in table 4.1. The deterministic part of the model is based on a system of Lotka-Volterra equations and describes changes in the mean frequencies of variants over all populations, to infer their fitness. Lotka-Volterra equations, together with their equivalent in evolutionary game theory, the replicator equation, constitute the limit process of the Wright-Fisher model for infinite population size, and for a symmetric game between two

strategies (Bomze, 1983; Imhof and Nowak, 2006). The stochastic part of the model is based on the Dirichlet distribution and describes the variance of variant frequencies between populations at each time-point. It is used to infer the strength of genetic drift. The Dirichlet distribution is frequently used to model the random fluctuations of a set of neutral markers in isolated populations derived from the same parental population (Kitakado et al., 2006) and changes in the frequencies of selectively neutral alleles under the Wright-Fisher model (Ewens, 2004).

TABLE 4.1 – **Parameters and state variables for the Lotka-Volterra Dirichlet-multinomial model.**

	Designation (unit) [reference value]
Observed variables	
$x_i^p(t)$	Sequence counts for variant i in population p at time t (<i>seq</i> ^a)
$x_{tot}^p(t)$	Total sequence counts in population p at time t (<i>seq</i>)
$n_{inf}(t)$	Number of infected hosts at time t (<i>host</i>)
$div(t)$	Mean variant diversity at time t (<i>variant</i>)
State variables	
$\lambda_i(t)$	Mean frequency of variant i at time t (no unit)
$\mathbf{X}^p(t)$	Random variable for sequence counts in population p at time t (<i>seq</i>)
$\Lambda^{\theta_{LV}}(t)$	Random variable for variant frequencies at time t (no unit)
Parameters of interest	
r_i	Relative intrinsic rate of increase of variant i (<i>generation</i> ⁻¹) ^b
$\bar{N}_e(t)$	Harmonic mean of effective population sizes at time t (individuals)
Fixed parameters	
μ	Point mutation rate per generation and per nucleotide (<i>generation</i> ⁻¹ · <i>nt</i> ⁻¹) [10 ⁻⁵] ^c
γ	Number of virus generations per day (<i>generation</i> · <i>day</i> ⁻¹) [1] ^d
λ^{inoc}	Vector of variant frequencies in the inoculum (no unit)
\mathbf{T}^{obs}	Vector of measurement dates (<i>day</i>) [(6, 10, 14, 20, 27, 34)]

^a The abbreviation "seq" is the number of virus cDNA sequences obtained.

^b The mean intrinsic rate of increase \bar{r} of all variants is one.

^c The value 10⁻⁵ is in the range of estimations for plant RNA viruses (Drake and Holland, 1999; Malpica et al., 2002; Tromas and Elena, 2010).

^d The number of virus generations per day, γ , was set to unity (Khelifa et al., 2010).

Data definition λ^{inoc} denotes the vector of the observed variant frequencies in the parental population — the inoculum used to inoculate all host plants in the real-life experiment (or to initialize the numerical experiment). At each measurement date t , $n_{inf}(t)$ hosts were effectively infected by the virus. We measured the vector $\mathbf{x}^p(t) = (x_1^p(t), \dots, x_{n_{var}}^p(t))$, with $x_i^p(t)$ the

number of obtained sequences of virus variant i ($1 \leq i \leq n_{var}$) in population p ($1 \leq p \leq n_{inf}(t)$) at measurement date $t \in \mathbf{T}^{obs}$. Thereafter, the t -notation will be used only for elements of \mathbf{T}^{obs} . The total number of sequences obtained from population p at time t is $x_{tot}^p(t) = \sum_{i=1}^{n_{var}} x_i^p(t)$. From this, the mean variant diversity at date t can be defined as $div(t)$, the mean number of variants with frequencies higher than 1% (i.e. such that $x_i^p(t)/x_{tot}^p(t) \geq 0.01$) over the $n_{inf}(t)$ population observed.

Parameters of interest The vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_{LV}, \boldsymbol{\theta}_{N_e})$ is the vector of all inferred parameters. $\boldsymbol{\theta}_{LV} = (r_i)$ is the vector of the intrinsic rate of increase r_i of each virus variant i . We assumed that the mean intrinsic rate of increase \bar{r} of all variants is one, as we are interested only in the relative intrinsic rates of increase. The selection coefficient of a variant i is usually computed as $s_i = r_i - 1$. $\boldsymbol{\theta}_{N_e} = (\bar{N}_e(t))$ is, over the various $t \in \mathbf{T}^{obs}$, the vector of parameters describing the harmonic mean of effective population sizes. We considered experimental design where samples are independent both between observation time-points and within each measurement date. Parameters can then be easily estimated by maximum-likelihood as detailed in the Materials and Methods section. The model and all the analyses presented here were implemented in R software version 3.0.2 (R Core Team, 2013).

Model equations The mechanistic-statistical model is based on deterministic Lotka-Volterra equations and stochastic Dirichlet-multinomial processes. It can be expressed as :

$$\left\{ \begin{array}{l} \frac{d\lambda_i}{dt} = \lambda_i \left(r_i - \sum_j r_j \lambda_j \right) + \sum_{j \neq i} \mu_{ij} (\lambda_j - \lambda_i) \quad (4.1) \\ \mathbf{X}^p(t) \sim DM(\boldsymbol{\lambda}^{\boldsymbol{\theta}_{LV}}(t), \boldsymbol{\theta}_{Dir}(t), x_{tot}^p(t)) \quad \text{with } \boldsymbol{\theta}_{Dir}(t) = \frac{\bar{N}_e(t)}{t} \quad (4.2) \\ \text{and } \boldsymbol{\lambda}^{\boldsymbol{\theta}_{LV}}(t) = (\lambda_i(t)), \quad i = 1, \dots, n_{var} \\ \bar{N}_e(t) = \frac{t}{\sum_{\delta=1}^d \frac{\gamma(\mathbf{T}^{obs}(\delta) - \mathbf{T}^{obs}(\delta - 1))}{\eta(\delta)}} \quad \text{with } \mathbf{T}^{obs}(0) = 0 \text{ and } T^{obs}(d) = t \quad (4.3) \end{array} \right.$$

Deterministic system The Lotka-Volterra equations (eq. 4.1) were assumed to be representative of the dynamics of the mean frequencies λ_i , defined for each variant i over all populations p across time. They combine (i) the logistic increase of variant i with parameter r_i , (ii) its competition with all other variants, and (iii) a mutation process, with μ_{ij} the mutation rate from variant j to variant i . Based on theoretical considerations (Sole et al., 1999) and their experimental counterpart (Fabre et al., 2012a), the coefficients of the competition exerted by variant j on variant i were taken as $\frac{r_j}{r_i}$, hence the $\lambda_i \sum_j r_j \lambda_j$ competition term.

Mutation rates were assumed to be symmetric and dependent only on the mutational distance between variants i and j , δ_{ij} (number of point mutations separating variant i from variant j), and the point mutation rate per generation and per nucleotide μ . If we define δ_{tot} as the total number of point mutations studied, and setting the number of virus generations per day to unity (Khelifa et al., 2010), we have $\mu_{ij} = \mu_{ji} = \mu^{\delta_{ij}}(1 - \mu)^{\delta_{tot} - \delta_{ij}}$. Hereafter, we will consider variants differing by at most $\delta_{tot} = 3$ point mutations.

Stochastic system The Dirichlet-multinomial distribution (eq. 4.2) was used to take into account random fluctuations from theoretical means λ_i due to (i) the heterogeneity of virus populations between hosts, and (ii) the noise introduced by the sampling process (Fabre et al., 2012a). The Dirichlet distribution accounted for the heterogeneity of the virus populations and the multinomial accounted for noise, as follows. Equation 4.2 assumes that $\mathbf{\Lambda}^{\theta_{LV}}(t)$, the vector of the true frequencies of all variants at time t in a given population, results from a Dirichlet distribution $\mathbf{\Lambda}^{\theta_{LV}}(t) \sim D(\boldsymbol{\lambda}^{\theta_{LV}}(t), \theta_{Dir}(t))$ of parameters $\boldsymbol{\lambda}^{\theta_{LV}}(t)$ and $\theta_{Dir}(t)$. Its mean is $E(\mathbf{\Lambda}^{\theta_{LV}}(t)) = \boldsymbol{\lambda}^{\theta_{LV}}(t)$, where $\boldsymbol{\lambda}^{\theta_{LV}}(t)$ corresponds to the vector of the mean variant frequencies (over all populations) at time t , determined by the Lotka-Volterra model with parameters $\boldsymbol{\theta}_{LV}$ and initial conditions $\boldsymbol{\lambda}^{inoc}$. Its variance, $Var(\mathbf{\Lambda}^{\theta_{LV}}(t)) = \frac{\boldsymbol{\lambda}^{\theta_{LV}}(t)(1 - \boldsymbol{\lambda}^{\theta_{LV}}(t))}{1 + \theta_{Dir}(t)}$, has two basic properties in common with genetic drift. First, $Var(\mathbf{\Lambda}^{\theta_{LV}}(t))$ decreases with increasing $\theta_{Dir}(t)$. Similarly, the differentiation of variant frequencies across a large set of completely isolated populations derived from the same parental population decreases with increasing effective population size (Charlesworth, 2009). Second, $\theta_{Dir}(t)$ has the same impact on all variants at a given time point, t , just like genetic drift (Balding, 2003).

The vector $\mathbf{\Lambda}^{\theta_{LV}}(t)$ is never observed in practice. In our case study, we observed, for each virus population, a sample of $x_{tot}^p(t) \simeq 3000$ individual viral sequences, using HTS to estimate $\mathbf{\Lambda}^{\theta_{LV}}(t)$. We therefore assumed that our observations $\mathbf{x}^p(t)$ of variant numbers resulted from a multinomial distribution conditional to $\mathbf{\Lambda}^{\theta_{LV}}(t)$, as $\mathbf{X}^p(t) | \mathbf{\Lambda}^{\theta_{LV}}(t) \sim M(\mathbf{\Lambda}^{\theta_{LV}}(t), x_{tot}^p(t))$. It follows that $\mathbf{X}^p(t)$ is issued from the Dirichlet compound multinomial distribution (eq. 4.2). The

Dirichlet-multinomial family is extensively used in population genetics, (e.g. Balding (2003); Balding and Nichols (1995); Beaumont and Balding (2004); Foll et al. (2014a); Kitakado et al. (2006)), and, more generally, for modeling uncertainty on relative frequencies (Corander et al., 2003; Shonkwiller and Hanley, 2003), and for the estimation of fixation indices (Balding and Nichols, 1995; Foll et al., 2014a; Kitakado et al., 2006).

4.2.2 Numeric experiment with FFPopSim

We performed several batches of simulations to assess the ability of the LVDM model to infer effective population sizes and selection coefficients accurately from experimental evolution studies in which isolated populations derived from the same parental population are sampled independently, at different time points. See supplementary information (Text S1) for the details on the numerical simulation. Briefly, we simulated the changes in variant frequencies under various selection and genetic drift regimes with FFPopSim (Zanini and Neher, 2012a), a Wright-Fisher model including mutation, selection and genetic drift for haploid individuals biallelic at multiple loci.

The simulations were designed to fit our real-life datasets where an evolve-and-resequence approach is used to estimate the frequencies of five virus variants at six time-points, in eight plants per sampling date. The results obtained nevertheless provided some more general insight. Overall, we generated 9000 numerical datasets in a full factorial design with five factors encompassing a wide range of selection and genetic drift regimes (Table 4.2). A numerical dataset gather all variants frequencies measured in a given host environment. The parameters $\theta = (\theta_{LV}, \theta_{N_e})$ were estimated with the LVDM model for each numerical dataset. We assessed the accuracy of the estimates by comparing directly (i) the true and estimated values of relative intrinsic rates of increase of each virus variants and (ii) true and estimated harmonic mean of effective population sizes at each sampling date.

A numerical dataset was simulated in two steps as follow. In step 1, we first defined the selection regime (factor σ_r) by the randomly sampling the fitness of each of the 5 variants considered (vector \mathbf{r}) and setting the mean fitness value to one. We secondly defined the genetic drift regime experienced by the virus population during the time course of the experiments (34 generations). We randomly sampled the effective population sizes at each generation (vector \mathbf{N}) according to four factors used to mimick the three main steps of host infection with a virus : (i) bottleneck at host inoculation, lasting one generation (factor N_e^{inoc}), (ii) genetic drift during multiplication in the inoculated organ, over five generations (factor N_e^{IO}) and (iii) genetic drift

during the colonization of all organs of the host, this stage being called ‘systemic infection’ and lasting 28 generations (factors N_e^S and σ_S). In step 2, we generated 48 independent FFPopSim simulations (given \mathbf{r} and \mathbf{N}), corresponding to the dynamics of populations in different hosts. We then sampled the composition of eight individual populations at each of the measurement dates corresponding to those used in the biological experiment, $\mathbf{T}^{obs} = (6, 10, 14, 20, 27, 34)$ dpi. At this step, the datasets generated were accepted if the frequency of at least two variants are $> 1\%$ in at least 100% (resp. 50%) of the populations at the first measurement date (resp. at any other date).

TABLE 4.2 – Parameters and factorial design for the numerical experiment.

Parameters	Designation (unit)	Levels and/or value
Factorial design		
N_e^{inoc}	Effective population size at inoculation (<i>seq</i>)	[10; 500]
N_e^{IO}	Effective population size in inoculated organ (<i>seq</i>)	[20; 5000]
\bar{N}_e^S	Harmonic mean of effective population sizes at the systemic level (<i>seq</i>)	[20; 5000]
σ_S	Standard deviation of instantaneous effec- tive population sizes at the systemic level	3 levels : $10^{0.05}$, $10^{0.5}$, 10^1
σ_r	Standard deviation of fitness values	3 levels : 0.025, 0.05, 0.1
Drawn		
\mathbf{r}	Vector of fitness values (<i>generation</i> ⁻¹)	
$N_e(\tau)$	Effective population size at generation τ (<i>seq</i>)	
Fixed		
n_{var}	Number of variants	5
ω^{inoc}	Vector of variant frequencies in the inoculum	(0.2, 0.2, 0.2, 0.2, 0.2)
\mathbf{T}^{obs}	Vector of measurement dates (<i>day</i>)	(6, 10, 14, 20, 27, 34)
x_{WF}^{tot}	Total sequence counts in one plant (<i>seq</i>)	3000
μ	Point mutation rate per generation and per nucleotide (<i>generation</i> ⁻¹ · <i>nt</i> ⁻¹) ^a	10^{-5}
γ	Number of virus generations per day (<i>generation</i> · <i>day</i> ⁻¹) ^b	1

^a The value 10^{-5} is in the range of estimations for plant RNA viruses (Drake and Holland, 1999; Malpica et al., 2002; Tromas and Elena, 2010).

^b The number of virus generations per day, γ , was set to unity (Khelifa et al., 2010).

4.2.3 Range of selection and genetic drift intensities explored in numerical simulations

As expected from the large parameter value space explored, the numerical experiment generated very contrasted selection (r_i parameters) and genetic drift (\bar{N}_e parameters) regimes (Fig. S4.1). The effective population sizes at inoculation (N_e^{inoc}), in the inoculated organ (N_e^{IO}) and during systemic infection (\bar{N}_e^S) were combined independently to explore a large range of effective population sizes $\bar{N}_e(t)$ at any measurement time. Low $\bar{N}_e(t)$ values were observed at all dates, with minimum values of about 20. The highest values of $\bar{N}_e(t)$ explored increased with time, from 1900 to 5400. Relative fitness values (r_i) ranged from 0.63 to 1.29 ($|s|$: 10% quantile = 0.0004, mean = 0.04, 90% quantile = 0.1), independently of effective population size (Fig. S4.1).

We thus obtained very different individual datasets with FFPopSim (Fig. 4.1). First, major differences in genetic drift regimes were observed both within and between datasets. For a given sampling date, variant frequencies could be similar (Fig. 4.1A,B) or different (Fig. 4.1C-E) between populations, due to the effective population size values. Moreover, for a given dataset, the independent sampling of effective population sizes N_e^{inoc} , N_e^{IO} and \bar{N}_e^S generated genetic drift regimes that varied over time. For example, in Fig. 4.1C, we can see strong similarities between populations at the first sampling date, but greater heterogeneity at subsequent dates. Second, selection regimes differed considerably between datasets. In particular, all variants remained present in all populations in some datasets (Fig. 4.1A,E). In others, some of the variants became extinct in all populations over time, due to strong selection, even in the presence of intermediate (Fig. 4.1C) or strong genetic drift (Fig. 4.1D). Finally, four of the five variants are essentially lost in Fig. 4.1B.

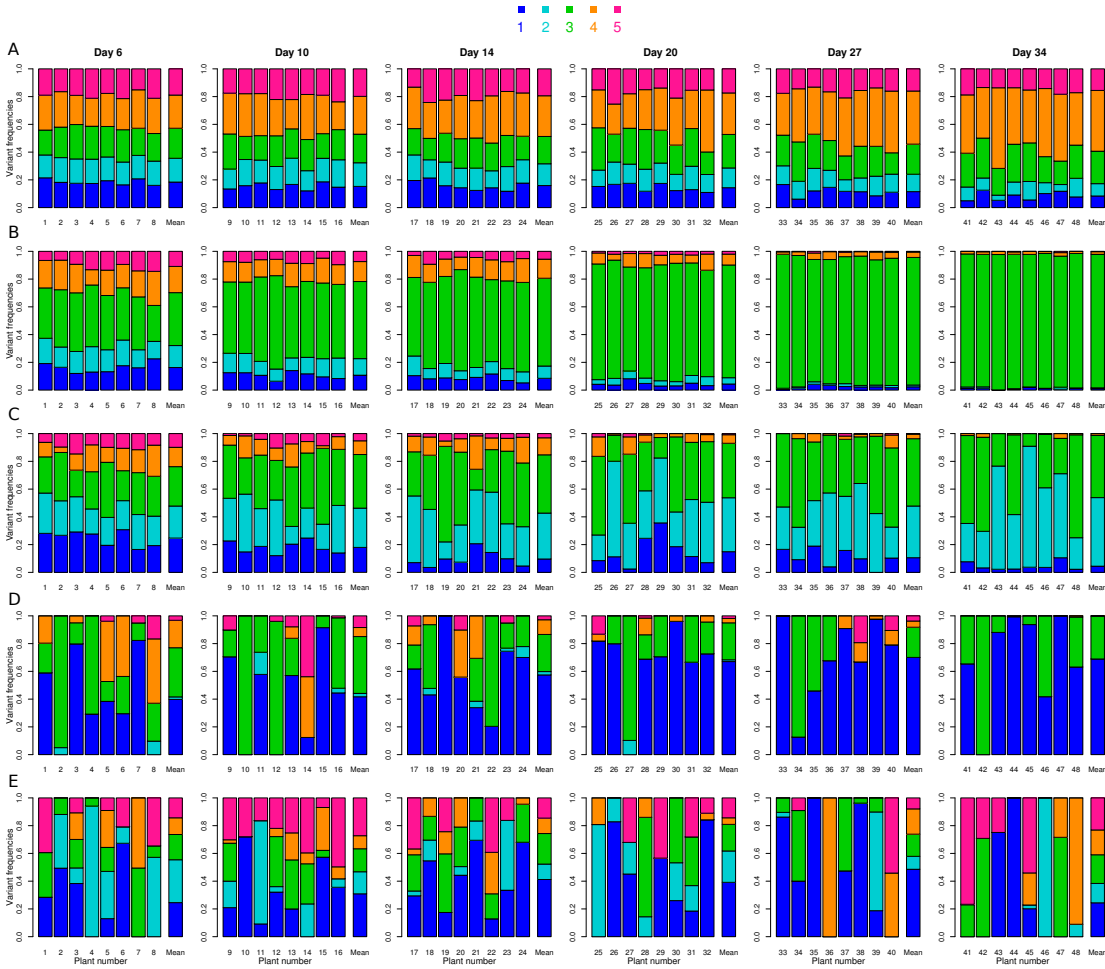


FIGURE 4.1 – **Contrasting datasets obtained in the numerical experiment.** For each dataset (series A to E), the composition of eight populations was observed at six measurement dates, from 6 to 34 days post inoculation, in independently sampled hosts. Within each bar plot, each bar represents the population composition in one host at one date, and the last bar shows the mean frequencies over these populations. One bar plot corresponds to one measurement date, indicated at the top of the column. A color code is used to distinguish between the five variants, indicated at the top of the figure. Parameter values are : series (A) $\mathbf{r}=(0.98, 0.98, 1.01, 1.03, 1)$, $N_e^{inoc}=450$, $N_e^{IO}=3116$, $\mathbf{N}_e^S=(1568, 2028, 2816, 3450, 2616, 2774)$; series (B) $\mathbf{r}=(0.97, 0.97, 1.12, 1, 0.93)$, $N_e^{inoc}=214$, $N_e^{IO}=2020$, $\mathbf{N}_e^S=(840, 1222, 1457, 1980, 2459, 2609)$; series (C) $\mathbf{r}=(1.01, 1.06, 1.07, 0.97, 0.88)$, $N_e^{inoc}=70$, $N_e^{IO}=459$, $\mathbf{N}_e^S=(239, 298, 295, 335, 380, 383)$; series (D) $\mathbf{r}=(1.14, 0.89, 1.08, 0.96, 0.93)$, $N_e^{inoc}=18$, $N_e^{IO}=25$, $\mathbf{N}_e^S=(23, 34, 41, 46, 54, 64)$; series (E) $\mathbf{r}=(1.04, 1, 0.99, 0.96, 1.01)$, $N_e^{inoc}=18$, $N_e^{IO}=22$, $\mathbf{N}_e^S=(21, 30, 40, 39, 41, 49)$.

4.2.4 Estimation of effective population size and variant fitness from simulated datasets

Effective population sizes $\bar{N}_e(t)$ and the intrinsic rates of increase of each variant r_i were inferred from the simulated datasets with the LVDM model. The estimated and true parameter values were compared to assess estimation accuracy.

Raw estimates of effective population size $\hat{N}_e(t)$ were satisfactory for the highest values

of $\bar{N}_e(t)$, but $\bar{N}_e(t)$ was underestimated when mean variant diversity at date t , $div(t)$, was $< n_{var}$ (Fig. S4.2). Underestimation occurred when at least one variant was lost (or nearly lost) in at least one population. It should be borne in mind that $div(t)$ is defined as the mean number of variants with frequencies higher than 1% over the eight populations observed at time-point t . Underestimation effects also increased in frequency over time, as demonstrated by the increasingly large swarm of blue points (for which $div(t) < n_{var}$) over black points (for which $div(t) = n_{var}$) in Fig. S4.2. Underestimation effects were also revealed by the slopes of the best-fit line between $\bar{N}_e(t)$ and $\hat{N}_e(t)$ and by the increasing (in absolute value) relative bias on $\bar{N}_e(t)$ (table 4.3).

TABLE 4.3 – Performance of the raw and bias-adjusted estimators of harmonic mean of effective population sizes $\bar{N}_e(t)$ and variant fitness r obtained with the Lotka-Volterra Dirichlet-multinomial model.

Parameter	R^2	Slope	Intercept	Mean bias [95% CI]
$\hat{N}_e(6 \text{ dpi})$	0.93	1.33	-0.83	-0.08 [-0.39; 0.09]
$\hat{N}_e(10 \text{ dpi})$	0.94	1.46	-1.18	-0.11 [-0.42; 0.07]
$\hat{N}_e(14 \text{ dpi})$	0.95	1.50	-1.34	-0.12 [-0.45; 0.06]
$\hat{N}_e(20 \text{ dpi})$	0.95	1.53	-1.45	-0.13 [-0.49; 0.07]
$\hat{N}_e(27 \text{ dpi})$	0.93	1.54	-1.51	-0.14 [-0.53; 0.08]
$\hat{N}_e(34 \text{ dpi})$	0.90	1.53	-1.56	-0.16 [-0.56; 0.09]
\hat{r}_i	0.92	0.79	0.21	$7 \cdot 10^{-4}$ [-0.03; 0.04]
$\hat{N}_e^{adj}(6 \text{ dpi})$	0.95	1.00	0.06	0.03 [-0.08; 0.13]
$\hat{N}_e^{adj}(10 \text{ dpi})$	0.95	1.03	0.01	0.03 [-0.07; 0.12]
$\hat{N}_e^{adj}(14 \text{ dpi})$	0.95	1.04	-0.04	0.03 [-0.07; 0.11]
$\hat{N}_e^{adj}(20 \text{ dpi})$	0.95	1.07	-0.11	0.02 [-0.08; 0.11]
$\hat{N}_e^{adj}(27 \text{ dpi})$	0.95	1.08	-0.14	0.02 [-0.1; 0.11]
$\hat{N}_e^{adj}(34 \text{ dpi})$	0.93	1.07	-0.12	0.01 [-0.13; 0.11]
\hat{r}_i^{adj}	0.94	0.98	0.02	$5 \cdot 10^{-5}$ [-0.03; 0.03]

For each parameter, the determination coefficient R^2 , the slope and the intercept of the best linear model fit between predicted and true values are given, together with the mean relative bias and its 95% confidence interval. Bias-adjusted estimators carry a 'adj' superscript, the other ones being raw estimators.

The raw estimator $\hat{N}_e(t)$ could be corrected to provide accurate estimates of $\bar{N}_e(t)$ over the whole range of $\bar{N}_e(t)$ values. This bias-adjusted estimator of $\bar{N}_e(t)$, called $\hat{N}_e^{adj}(t)$, was defined as follows,

$$\begin{cases} \hat{N}_e^{adj}(t) = \hat{N}_e(t) + N_e^C(t) & \text{if } div(t) < n_{var} \end{cases} \quad (4.4)$$

$$\begin{cases} \hat{N}_e^{adj}(t) = \hat{N}_e(t) & \text{if } div(t) = n_{var} \end{cases} \quad (4.5)$$

Only the raw estimates for which $div(t) < n_{var}$ were corrected by adding $N_e^C(t)$, defined as the effective population size expected from a neutral Wright-Fisher model with $div(t)$ alleles after t generations. According to Karlin (1968) (page 462), this size is $N_e^{NWF}(t) = [1 - (div(t)/n_{var})^{(1/t)}]^{-1}$ for a haploid population. The best correction was obtained with the lowest value between N_e^{NWF} and $t \times n_{var}$, leading to $N_e^C(t) = \min\{N_e^{NWF}, t \times n_{var}\}$. $\hat{N}_e^{adj}(t)$ provided very accurate estimates of the true effective population size $\bar{N}_e(t)$, with R^2 determination coefficients ranging from 0.93 to 0.95, and the best-fit lines between $\bar{N}_e(t)$ and $\hat{N}_e^{adj}(t)$ were very close to the first bisector ($1.0 < \text{slope} < 1.07$, $-0.13 < \text{intercept} < 0.06$, Fig. 4.2A-E and table 4.3). This correction brought the mean relative bias on $\bar{N}_e(t)$ closer to zero, and narrowed its 95% confidence interval.

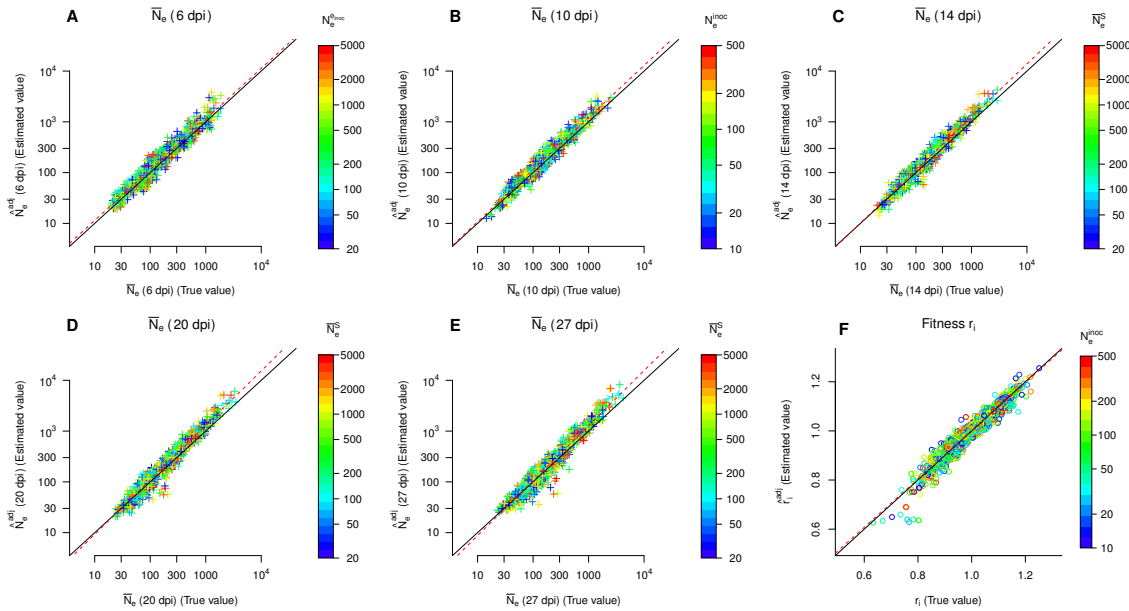


FIGURE 4.2 – Inference of the harmonic mean of effective population sizes \bar{N}_e and variant fitness r with the Lotka-Volterra Dirichlet-multinomial model from datasets simulated in the numerical experiments. (A-E) Correlation between the true \bar{N}_e (x-axis) and estimated \hat{N}_e^{adj} (y-axis) at 6, 10, 14, 20 and 27 days post-inoculation (dpi). (F) Correlation between true r_i (x-axis) and estimated \hat{r}_i^{adj} (y-axis) (intrinsic rates of increase of variants). In all panels, the black line is the first bisector and the red dashed line is the best linear model fit. For clarity, only 1/3 of the data (randomly drawn) are represented.

Raw estimates of the intrinsic rates of increase \hat{r}_i were satisfactory, with a R^2 of the best-fit line of 0.92 (Fig. S4.2). However, the slope of the best-fit line (0.79) resulted in a slight overestimation (when $r_i > 1$) or underestimation (when $r_i < 1$). This effect could be corrected with the following bias-adjusted estimator, $\hat{r}_i^{adj} = \hat{r}_i + \frac{(\hat{r}_i - 1)}{\sqrt{\hat{\theta}_{Dir}(6 \text{ dpi}) + 1}}$. The magnitude of correction is inversely proportional to the denominator of the standard deviation of the Dirichlet distribution of variant frequencies at 6 dpi. A slight increase in the determination coefficient

($R^2 = 0.94$, table 4.3) yielded a highly satisfactory correction for the slope (0.98) and the intercept (0.02) of the best-fit line between r_i and \hat{r}_i^{adj} and for the relative bias (Fig. 4.2F, table 4.3).

In a nutshell, from these numerical simulations with known parameter values, we can conclude that the proposed modeling framework provides accurate estimates of the intrinsic rates of increase r_i of each variant i , and thus of their selection coefficient, together with the dynamics of the effective population sizes $\bar{N}_e(t)$ during the time course of the experiment.

4.2.5 Observed dynamics of variant frequencies in real-life experiments with plant virus populations

In the real-life experiment, as in the numerical experiment, the frequencies of the five virus (PVY) variants were observed in completely isolated populations during the course of the infection, in 15 different host environments. For each of these 15 pepper genotypes, 48 plants were inoculated with the same equimolar mixture of the five variants, and the frequencies of the virus variants were observed in eight plants at each of six sampling dates, from 6 to 34 days post-inoculation.

In a few cases, no infection occurred in the inoculated plant, due to an extreme bottleneck at inoculation, resulting in the extinction of the virus population, or because systemic infection of the plant was delayed until after the period during which measurements were made (from 10 to 34 dpi). Cases of non-infection were observed in four doubled haploid (DH) lines (e.g. DH lines 219 and 2321, Fig. 4.3). The frequencies of non-infected plants were highest at 10 and 14 dpi, the first two dates on which systemically infected leaves were analyzed. An absence of infection was observed in a mean of 3.5 plants 10 dpi and in a mean of two plants 14 dpi, in the four DH lines concerned.

The virus populations present in all infected plants and in the common inoculum were analyzed by HTS, to determine the frequencies of the five PVY variants. During sequence analysis, we looked for all eight possible variants, given the three codon positions of interest in the VPg cistron, to ensure that none appeared by mutation or recombination, with possibly significant effects on PVY populations, compromising parameter estimation. All new variants remained below a reasonable threshold (4%). They were not, therefore, considered in subsequent analyses.

Inoculum analysis confirmed that all variants were present in roughly equimolar proportions,

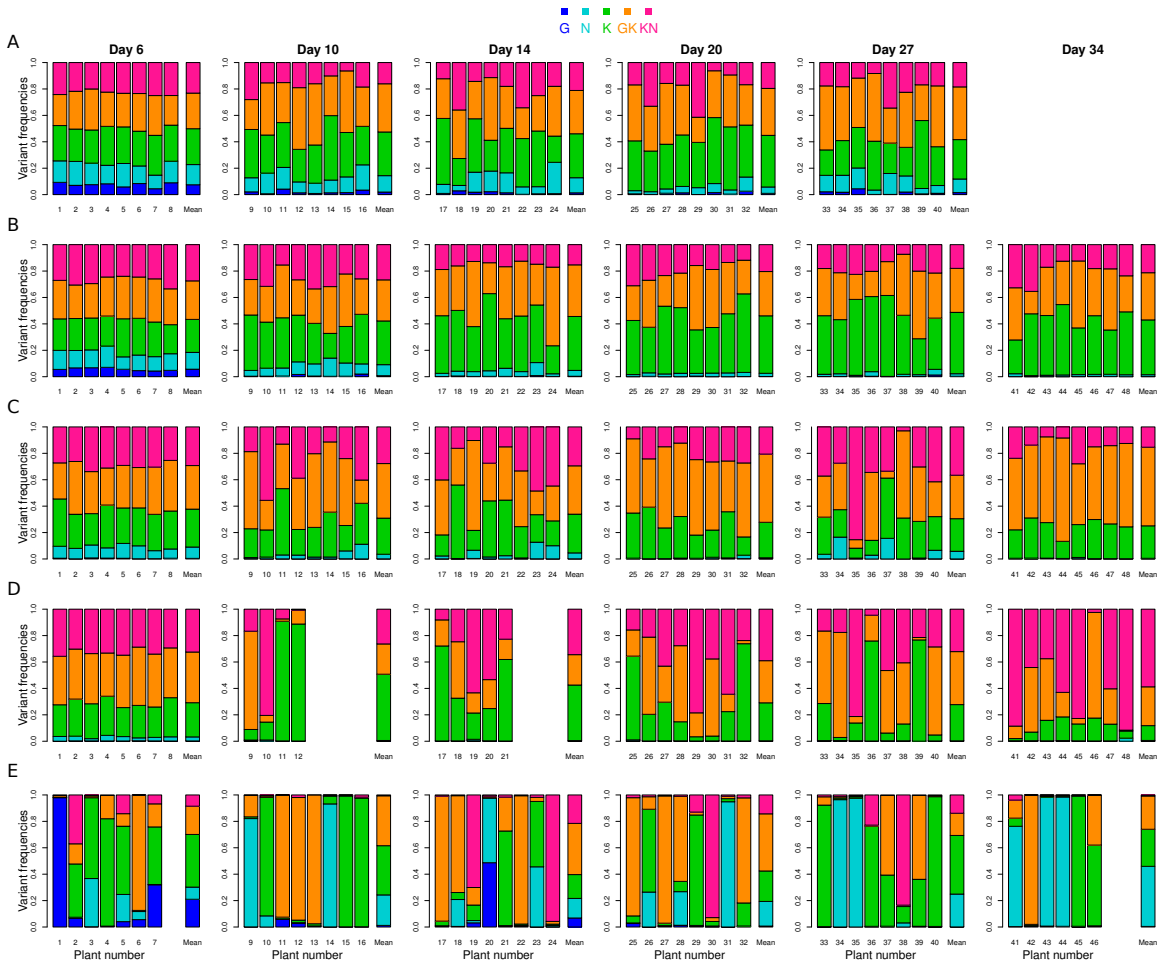


FIGURE 4.3 – **Five contrasting datasets obtained in the biological experiment.** One line of bar plots represents the dynamics of virus variants in one DH line : (A) 240, (B) 2430, (C) 2344, (D) 2321 and (E) 219. Within each bar plot, the frequencies of the five variants (see the top of the figure for the color code) in each infected plant sample are represented by one bar. The last bar indicates the mean viral composition over the infected plants. One bar plot corresponds to one measurement date indicated on the top of each column of barplots. The five DH lines displayed contrasting virus variant dynamics, consistent with contrasting patterns of selection and genetic drift. We could not sample plants of DH line 240 (A) 34 days post-inoculation because severe necrosis symptoms invading the stem led to all plants death at that late date.

with 22.6% of variant G, 17.5% of N, 20.6% of K, 17.1% of GK and 22.2% of KN.

The raw data for variant frequency dynamics provided diverse patterns, differing considerably between host environments (Fig. 4.3 and Fig. S4.3 and S4.4). As in the numerical experiment (Fig. 4.1), variant frequencies were similar between populations on the same date, in some host environments, consistent with weak genetic drift (e.g. DH lines 240 and 2430, Fig. 4.3A & B), whereas they differed in other host environments (e.g. DH lines 2321 and 219, Fig. 4.3D & E). Furthermore, the heterogeneity of variant frequencies between the eight plants analyzed fluctuated between dates, probably due to changes in effective population size during the course of infection (e.g. DH line 2344, Fig. 4.3C). Selection regimes were also variable. In

some DH lines, all variants remained present at late infection dates (e.g. DH line 240, Fig. 4.3A), whereas in others, one variant (e.g. DH line 219, Fig. 4.3E), or up to two variants (e.g. DH lines 2430, 2344, 2321, Fig. 4.3B-D)) became extinct. The four pepper genotypes for which some plants remained uninfected were also characterized by the highest heterogeneity in variant frequencies. Overall, the patterns encountered in the real-life datasets were similar to some of those obtained for the simulated datasets. However, the simulated datasets also included a number of datasets with more extreme patterns of selection and genetic drift regimes.

4.2.6 Estimation of effective population sizes and variant fitness from real-life datasets

We used the LVDM model to estimate $\bar{N}_e(t)$ and r_i for each DH line, to determine the quantitative evolutionary forces underlying these qualitative observations. By contrast to conditions in the numerical experiment, the evolutionary parameters underlying the true demogenetic dynamics of virus populations in their hosts are unknown. However, goodness-of-fit was frequently satisfactory for the LVDM model. Most R^2 values between the predicted and observed mean variant frequencies grouped around the median of the distribution of R^2 values for the FFPopSim numerical experiment (R^2 close to 0.89), with the exception of those for four DH lines : 2321 and 2400 with $R^2 = 0.76$, and 219 and 2256 with $R^2 = 0.44$ and 0.54, respectively (Fig. 4.4). All R^2 values corresponding to the biological experiment were nevertheless within the upper 95% of values of the R^2 distribution (5%-quantile at $R^2 = 0.33$). The R^2 values between predicted and observed standard deviation of variant frequencies were also grouped around the median of the R^2 distribution of the numerical experiment, lying within the upper 80% of the values of the R^2 distribution (20%-quantile at $R^2 = 0.44$). Selection coefficients took values from 0.008 to 0.15 (hence the range was included in the 95% confidence interval of the numerical experiment : 2.5% quantile = 0.001 and 97.5% quantile = 0.15).

Estimates of effective population size and intrinsic rates of increase are presented in Fig. 4.5. The fitness ranks of the PVY variants were globally the same in all DH lines. Variant G was the weakest, followed by variants N, KN, K and GK (Fig. 4.5A). This ordering was strictly conserved in the 13 DH lines in which the fitness differences between PVY variants were largest. Some differences in order were observed in the two DH lines with globally smaller fitness differences between variants. For DH line 219, variant KN had a lower fitness than variant N, and for DH line 2321, variant KN was the fittest. The fitness difference between the

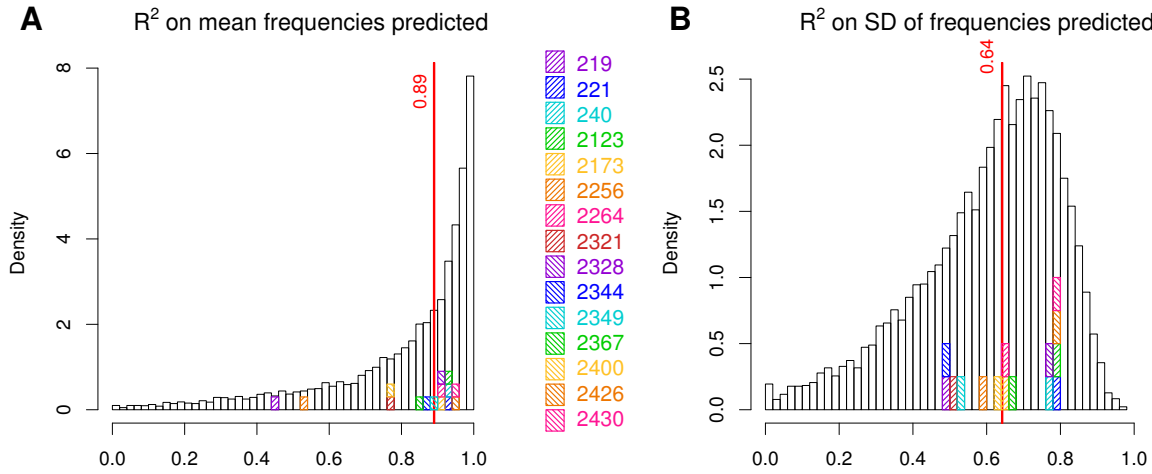


FIGURE 4.4 – **Goodness of fit of the Lotka-Volterra Dirichlet-multinomial model for the numerical and real-life biological experiments.** Distribution of Pearson’s R^2 values between adjusted and observed mean variant frequencies (A) and standard deviation of variant frequencies (B) for each dataset of the numerical experiment. Red lines indicate the median. Shaded rectangles locate the R^2 values of each host genotype (DH line) of the biological experiment.

weakest and the fittest variant, based on the bias-adjusted estimations \hat{r}_i^{adj} , ranged from 0.09 for DH line 219, to 0.23 for DH line 2430.

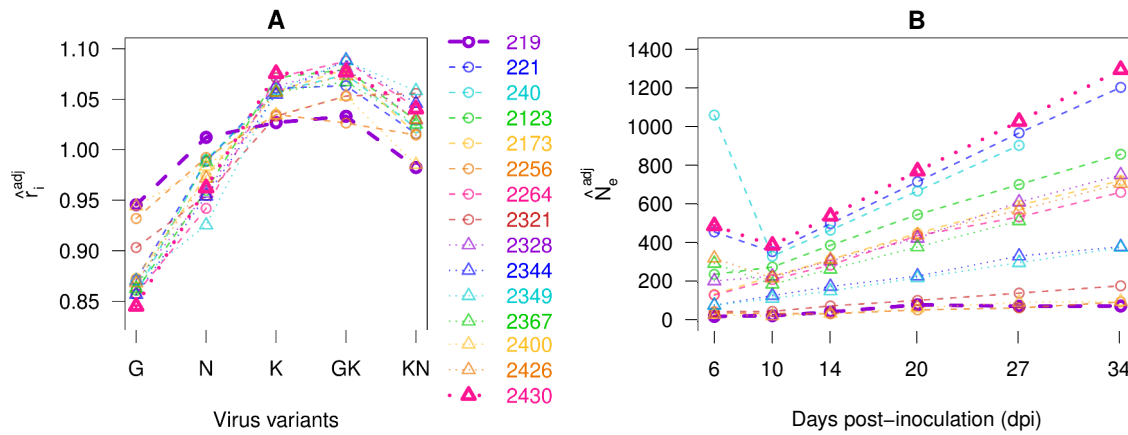


FIGURE 4.5 – **Fitness of virus variants and effective population size estimates for the 15 plant genotypes (DH lines).** (A) Bias-adjusted estimations of intrinsic rates of increase \hat{r}_i^{adj} for each variant i and for all DH lines. (B) Bias-adjusted harmonic mean of effective population size estimates $\hat{N}_e^{adj}(t)$ at each measurement date and for all DH lines. For readability, dots for the same DH line are connected by dashed or dotted lines. Two highly contrasting DH lines, 219 and 2430, are shown in bold.

We further estimated the dynamics of effective population size over the time course of the experiment, with one estimate derived for each sampling date. At the first measurement date, considerable variability was observed (Fig. 4.5B), $\hat{N}_e^{adj}(t)$ ranging from 25 for DH line 219 to 1059 for DH line 240). This was not surprising given that we chose the DH lines on the basis of the density of primary infection foci in inoculated organs (Fig. S4.7). A much narrower range of

$\hat{N}_e^{adj}(t)$ was observed at 10 dpi, the first date on which systemic infection was observed, across all host environments ($28 \leq \hat{N}_e^{adj}(10 \text{ dpi}) \leq 390$). From 6 to 10 dpi, $\hat{N}_e^{adj}(t)$ decreased in six DH lines, remained approximately constant in five DH lines and increased in four DH lines. By contrast, from 10 to 34 dpi, effective population size increased almost linearly in all DH lines except DH line 219, in which it decreased slightly after 20 dpi. The DH lines with the lowest values of $\hat{N}_e^{adj}(t)$ (i.e. lines 219, 2256, 2321 and 2400) also displayed the smallest differences in fitness between the weakest and the fittest variants (Fig. 4.5).

4.3 Discussion

Advances in sequencing technologies have revolutionized the collection of time series data, elucidating the evolutionary histories of many species with unprecedented accuracy. This study, for example, is the first to present findings suggesting such strong variability in the selection and genetic drift regimes experienced by plant viruses in closely related host genotypes (Fig. 4.3 and Fig. S4.3 and S4.4). This new type of dataset paves the way for the estimation of population genetics parameters influencing the fate of variants of special interest (e.g. with drug resistance). However, analytical methods must be developed for efficient estimation of the main evolutionary forces over their whole range of variation, and this step remains challenging.

4.3.1 A method for the joint estimation of genetic drift and selection over a large range of intensities from experimental evolution data

We present a mechanistic-statistical model based on deterministic Lotka-Volterra equations and stochastic Dirichlet-multinomial processes for the joint estimation of selection and genetic drift for small effective population sizes ($N_e \ll 1000$) and a wide range of selection coefficients (weak ($|s| \simeq 0.01$) or strong ($|s| \simeq 0.1$) selection). We assessed the efficacy of the LVDM model on numerical datasets for the evolution of a haploid, asexual population. The datasets were generated under the Wright-Fisher model, over wide ranges of initial number of founders ($N_e^{inoc} \in [10, 5000]$), subsequent population dynamics over several tens of generations ($N_e^{IO} \times \bar{N}_e^S \in [20, 5000] \times [20, 5000]$) and selection coefficient ($|s|$: 10% quantile = 0.0004, mean = 0.04, 90% quantile = 0.1) (Fig. S4.1). After adjustment for bias, the accuracy and precision of the estimates of s for each competing genotype and of the dynamics of \bar{N}_e over the time course of

the experiment were highly satisfactory ($R^2 \in [0.93, 0.95]$, slopes $\in [0.98, 1.07]$ and intercepts $\in [-0.14, 0.06]$ for the best linear fit between true and estimated parameters). These results were obtained for the evolution of isolated populations derived from the same parental population. The experimental data used here described the competition between five virus variants, the frequencies of which were determined by sequencing at six time-points (from generation 6 to generation 34) in eight replicates per date. Satisfactory estimates were also obtained when we considered data for competition between only two virus variants (Fig. S4.5), suggesting that the LVDM method and subsequent bias adjustment are not limited to the experimental design presented.

Recent reviews (Foll et al., 2014c; Malaspinas, 2016) have highlighted the small number of methods used to date for the joint estimation of selection and genetic drift over their whole range of variation, particularly in cases of small effective population size ($N_e \ll 1000$) and strong selection coefficient ($|s| \simeq 0.1$). This lack of methods reflects the fact that these conditions do not fulfill the hypothesis underlying most approximations of the Wright-Fisher model. The classical approximation, with a standard diffusion process, requires both selection and genetic drift to be weak (Malaspinas, 2016). Approximations based on Gaussian diffusion require the stochastic effects of genetic drift to decrease more rapidly than those of selection (Malaspinas, 2016). A major step forward was made by Foll et al. (2014b,c), who described a two-step method. Effective population size is first estimated with Fs' , an unbiased estimator of N_e proposed by Jorde and Ryman (2007) (Foll et al., 2014b). Fs' is a F -statistic calculating the variance of gene frequency changes between two time points, and taking the corresponding sample sizes into account to adjust for sampling error. In the second step, selection coefficients are inferred by ABC, through simulation of the Wright-Fisher model with an initial allele frequency and sample sizes matching those observed. The method requires at least 90% of the genetic markers studied to be under neutral selection. This assumption is not valid for all species, particularly pathogens with small genomes, such as viruses. For example, only 22.7% of 66 randomly chosen mutations in the *Tobacco etch virus* (TEV, genus *Potyvirus*) genome, a plant RNA virus, were found to be neutral (Carrasco et al., 2007). Similar results have been obtained for bacteria (e.g. Kassen and Bataillon (2006)). By contrast, the LVDM model does not assume that most of the loci studied are under neutral selection. Lacerda and Seoighe (2014) recently developed a new approximation of the Wright-Fisher model. Their method does not require neutral markers and can deal with selection of any intensity. Their method provided satisfactory estimates of both N_e and s (estimated for a single locus) for a relatively small effective population size of 1000 individuals and values of s up to 0.5. They did not test the performance of their method

for $N_e \ll 1000$. By comparison, the LVDM model was effective for much lower N_e values, in the range of a few tens of individuals, and for inferring the time course of \bar{N}_e over a few tens of generations. However, although the range of selection coefficients s included cases of strong selection ($|s| \simeq 0.1$, as defined by Malaspinas (2016)), none of the simulation experiments included values as high as 0.5. The inference of such very strong selection coefficients with the LVDM model may be possible, provided that the first generations are sampled more densely, typically everyday after inoculation in our set-up. Lacerda and Seoighe (2014), for example, used samples taken at each generation for 20 generations.

The LVDM model, rather than being based on an approximation of the Wright-Fisher model, tackles a problem from one scientific field — population genetics — with the approaches of another — population dynamics. These two fields are very closely related, particularly as concerns the use of Lotka-Volterra systems and the Wright-Fisher model (Bomze, 1983; Imhof and Nowak, 2006). In the LVDM model, selection is evaluated from the mean variant frequency trajectories with deterministic Lotka-Volterra equations for competition between variants including selection and mutation. Genetic drift is evaluated at each measurement date, from the variance of variant frequencies between the replicated populations observed on the date concerned, with a stochastic Dirichlet-multinomial process with the expected frequencies given by the Lotka-Volterra equations. The underlying rationale is that selection and genetic drift have different effects on variant dynamics. Their respective signatures are revealed by analyzing a set of independent trajectories of alleles. Even for populations of small effective size, for which genetic drift and selection have confounding effects on the fate of variants (Fig. 4.1), a moderate number of replicates contain enough information to disentangle the two mechanisms. Here, we estimated four selection coefficients and six effective population sizes (*i.e.* 10 parameters) with 48 samples (6 sampling dates \times 8 replicates). The LVDM model can also efficiently estimate the selection coefficient of each variant and a single effective population size, if eight replicates are available for a single sampling date (*i.e.* 5 parameters and 8 replicates). However, this required more drastic criteria for accepting the datasets than in the main experiment (Fig. S4.6). In the study closest to ours, Lacerda and Seoighe (2014) showed that the efficient estimation of two parameters (N_e and s) required 20 samples (one sample at each generation), but for much stronger selection coefficients ($s = 0.2$ or $s = 0.5$) and higher N_e (1000). Experimental replicates, that is, independent sets of allele trajectories, are generally easy to establish with experimental evolution designs. The use of such independent time-sampled data has several advantages. First, maximum-likelihood estimations are easy to handle, fast and efficient. Only a few seconds are

required with the 'nlminb' optimization routines of R to estimate the 10 parameters of interest in our case studies. By contrast, handling time dependence between observations can substantially complicate and delay likelihood-based estimation approaches. Second it is also more convenient, because if, for example, a variant becomes fixed in a population, then subsequent observations in the same population would not be informative. Finally, sampling the same population several times involves the removal of individuals, with possible consequences for the future evolutionary dynamics of the population, biasing estimates.

Several substantial improvements could be made to the LVDM model. We found that raw estimators were biased as soon as one of the variants became either fixed or extinct in at least one of the observed populations (*i.e.* $div(t) < n_{var}$, Fig. S4.2)). Variant loss is known to bias effective population size estimates (Luikart et al., 1999; Richards and Leberg, 1996). Similarly, some recently developed methods (e.g. Terhorst et al. (2015), reviewed by Malaspinas (2016)) have assumed that the frequency of the alleles under selection remain far from the thresholds associated with either fixation or extinction. In this paper, we employed a heuristic approach based on the work of Karlin (1968), in particular, to derive bias-adjusted estimators of \bar{N}_e and s that efficiently correct the estimates of \bar{N}_e for its lowest values. Further theoretical studies of this aspect would be valuable and would be likely to benefit from a recent study by Tataru et al. (2015), given that the Dirichlet distribution is the multivariate generalization of the beta distribution. These authors introduced an extension of the beta distribution, 'the beta with spikes', that explicitly models the loss and fixation probabilities as two spikes at the boundaries. The LVDM model explicitly takes the sampling noise into account, through the multinomial distribution hypothesis. This noise arises because variant frequencies are determined from finite counts of virus sequences. However, high-throughput sequencing also introduces sequencing errors, albeit at a low rate of about 1 substitution per 400 bases for MiSeq technology (Laehnemann et al., 2016), which were not explicitly accounted for in our framework. Several models have been proposed to disentangle true genetic variation from technical artifacts (Laehnemann et al., 2016) and these models could be integrated into the LVDM model, via a hierarchical Bayesian modeling framework (Clark, 2005), for example. Finally, the LVDM model ignores recombination and considers unlinked alleles. This is relevant for our case study, as the region of the viral genome analyzed is small, the nucleotide positions differentiating the variants being located only a few codons apart. In this situation, recombination can be considered negligible (Rhodes et al., 2005). Moreover, the small number of generations considered decreases the probability of recombination (Foll et al., 2014c). However, recombination is likely to become important if nucleotide positions distributed over longer genome regions are considered.

4.3.2 Plant genotypes modulate genetic drift and selection within virus populations.

The biological experiment presented in this article on the PVY- pepper pathosystem yielded informative results. The considerable diversity of selection and genetic drift regimes experienced by plant viruses in closely related host genotypes in this study has not, to our knowledge, been observed before (Fig. 4.3 and Fig. S4.3 and S4.4). On the host side, the experiment involved 15 DH lines of pepper, all carrying the major resistance gene *pvr2*³ and differing by several genes with quantitative resistance effects (Quenouille et al., 2013). On the pathogen side, we used five virus variants : the G and N variants displayed weaker adaptation to *pvr2*³, whereas the K, GK and KN variants displayed stronger adaptation. The ranking of the selection coefficients of the five variants, and thus of their fitness landscape, was mostly identical in the 15 plant genotypes. We were therefore unable to identify any host genotype, among those tested, with a contrasting fitness landscape counterselecting against the virus variants best adapted to *pvr2*³. However, the gradient of the fitness landscape, as estimated by the difference in fitness between the weakest and the fittest variants, varied somewhat between host genotypes. The speed with which the best-adapted variants are selected can therefore be modulated by carefully choosing the plant genotypes grown. We also inferred the genetic drift landscapes experienced by the viruses in the 15 host environments during the time course of the experiment. Genetic drift landscapes were highly variable between plant genotypes, revealing an unprecedented level of variability between closely related host genotypes. Our estimates of \bar{N}_e ranged between 28 and 390 just after the colonization of apical leaves at 10 dpi, and from 25 to 1059 in the inoculated leaves four days previously (at 6 dpi). We thus demonstrated a general bottleneck for all DH lines during the systemic movement of the virus. This result sheds new light on the relative importance of within-host bottlenecks experienced by virus populations, notably compared to a recent article by Zwart et al. (2011), who reported that the N_e of TEV in the first systemically infected leaf of tobacco plants was determined largely by the viral load of the inoculum. They then hypothesized that genetic drift occurred mostly during the inoculation process. Previous estimations of N_e for viruses did not focus on whole-plant scale N_e dynamics as in this study. Instead, they considered the multiplicity of infection (MOI) during cell-to-cell movement or N_e during the colonization of apical leaves (for a comprehensive review, see Gutiérrez et al. (2012a)). Direct comparison between these studies is not, therefore, appropriate. Gutiérrez et al. (2015) recently showed that *turnip mosaic virus* (genus *Potyvirus*) infections are characterized by a very low MOI ($\simeq 1$) when cells are infected with virus particles moving in

the plant vasculature and a much higher MOI ($\simeq 30$) during subsequent cell-to-cell movement in the mesophyll. The general picture that emerges when we consider both these MOI patterns and plant growth dynamics, is consistent with our observations. Lower N_e values were observed during the initial colonization of apical leaves, as the strongest bottleneck occurred during long-distance virus movement from the inoculated leaves to establish systemically infected leaves, at a stage of plant growth at which there were few leaves and even fewer infected leaves. \bar{N}_e tends to increase with time, because (i) increasing numbers of leaves are infected and behave as virus sources as the plant grows and (ii) leaf area index increases, probably increasing the relative proportion of cell-to-cell, as opposed to long-distance, virus movement.

The plants of each DH line were genetically identical, and experimental conditions were set up so as to prevent differences in environmental effects between DH lines. We were therefore able to show that the differences in genetic drift and selection observed between DH lines were controlled by plant genes. The heritability of the traits (i) intrinsic rates of increase and (ii) effective population size was estimated as $h^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_e^2/n)$, where σ_G^2 corresponds to the genotypic variance, σ_e^2 to the phenotypic variance and n to the number of replicates (Gallais, 1990). We artificially created two replicates for each DH line by dividing the dataset in two, with each half corresponding to a random choice of four of the eight plants for each sampling date. High heritability estimates were obtained for both the intrinsic rates of increase ($h^2 = 0.79$) and for effective population size ($h^2 = 0.94$). Furthermore, these estimates are almost certainly underestimates, because half the plants were used to estimate the two traits.

As \bar{N}_e is a heritable plant trait, our findings pave the way for the breeding of plant cultivars to expose the viruses that infect them to greater genetic drift. This would provide a double-edged sword against viruses. First, in asexual populations, genetic drift favors the accumulation of deleterious mutations, decreasing viral fitness (Muller's ratchet) (de la Iglesia and Elena, 2007). Second, genetic drift decreases the fixation probability of beneficial mutations, such as those overcoming plant resistance genes (Patwa and Wahl, 2008). Breeding for increased genetic drift in virus populations would constitute a novel approach to manage the durability of resistance to plant viruses in agricultural landscapes (Fabre et al., 2012b, 2015). These findings raise questions about how within-host fitness and genetic drift effects act together with between-host fitness, genetic drift and epidemic effects to induce the emergence of a new pathogen variant. Despite being much simpler than the underlying mechanisms involved in the infection of highly structured hosts, the simple Wright-Fisher model accurately captures the major processes driving the within-host demogenetic dynamics of a set of virus variants. It can thus serve as a valuable cornerstone for linking the within- and between-host scales of disease

dynamics.

4.4 Materials and Methods

4.4.1 Biological experiment

In the biological experiment, we used plant RNA viruses to study the evolutionary forces of genetic drift and selection experimentally.

Plant and virus materials We used 15 doubled haploid (DH) lines of pepper (*Capsicum annuum*; family Solanaceae) as different host environments. All the plants of a given genotype were thus genetically identical. These lines were chosen for this study on the basis of quantitative differences in three previously measured factors, so as to generate different intensities of genetic drift and selection acting on virus populations : (i) relative viral accumulation, (ii) the resistance-breakdown (RB) frequency (Quenouille et al., 2013) and (iii) the number of primary infection foci after mechanical inoculation (see Fig. S4.7).

All plants were mechanically inoculated with the same initial virus population comprising an equimolar mixture, based on quantitative double antibody sandwich - enzyme linked immunosorbent assay (DAS-ELISA), of the five *Potato Virus Y* (PVY; genus *Potyvirus*, family Potyviridae) variants G, N, K, GK and KN (Fabre et al., 2012a). Single-letter and double-letter names correspond to single and double mutants, respectively, of the infectious clone SON41p (Fig. S4.8). Three mutations ($\delta_{tot} = 3$) located close together in the genome differentiate the five variants, and these mutations are named after the amino-acid substitutions observed at positions 101 for the S (serine) to G (glycine) substitution, 115 for the T (threonine) to K (lysine) substitution, and 119 for the D (aspartic acid) to N (asparagine) substitution, of the VPg (viral protein genome-linked) protein. The G and N variants displayed a low level of adaptation to the chosen pepper genotypes, whereas variants K, GK and KN displayed higher levels of adaptation (Quenouille et al., 2013).

Experimental set-up, plant sampling and High-Throughput Sequencing. For each pepper genotype, 48 plants were arranged in randomized blocks, to minimize environmental effects. The first true leaf of each plant was inoculated 29 days after sowing. We then analyzed eight plants per DH line at 6, 10, 14, 20, 27 and 34 days post-inoculation (dpi) (Fig. S4.8). The inoculated leaf was sampled at 6 dpi, and, on subsequent sampling dates, three uninoculated leaves, corresponding to the three youngest unfolded leaves, were sampled and pooled together.

As the plants were removed after sampling, all the virus population datasets obtained from each sample were independent. Each leaf sample was ground in four volumes of 0.03 M phosphate buffer (pH 7.0) supplemented with 2% (w/v) diethyldithiocarbamate, as previously described in Fabre et al. (2012a).

Total RNA was purified from individual plant samples with the Tri-Reagent kit (Sigma-Aldrich). We then amplified a 104-nucleotide region encompassing the polymorphic region of the PVY VPg cistron by reverse transcription-polymerase chain reaction (RT-PCR) with tagged primers. Eight differently tagged primers were used, corresponding to the eight different plant replicas of the same plant genotype for each sampling date (Table S4.1). Amplified DNAs corresponding to the eight plant replicas were pooled together on the basis of their intensity on 2% electrophoresis gels and 2×150 bp paired-end libraries with multiplex adapters were prepared. HTS was performed with a MiSeq Illumina sequencer at the Genomic Platform of INRA Toulouse (MiSeq Reagent Kit V2 500 cycles). In the initial sequence analysis, we used FLASH software to obtain the consensus sequence from reads 1 and 2 with a minimum overlap length between the two reads of 63, a maximum overlap length of 153 and a maximum allowed ratio between the number of mismatched base pairs and the overlap length of 0.2 (Magoç and Salzberg, 2011). Then, the sequences were identified by adapter and by tag to attribute a sampling date and a plant to each of them. Sequences of each population were filtered with the help of the 'agrep' function in R software, to retain only the sequences showing less than 20% differences with the infectious clone SON41p. Finally, the three codons of interest were analyzed to identify PVY variants and determine the composition of the virus population in each sample (observed variable $x^p(t)$). After sequence filtering, we had 374 to 14141 sequences per sample (observed variable $x_{tot}^p(t)$), with a mean of 3295 sequences per sample. The raw data issued from this analysis (*i.e.* Number of sequences and composition of the virus population in each of the XXX samples analysed) are available in Table S4.2.

4.4.2 Maximum-likelihood estimation

Let \mathbf{x} denote the vector of all sequence counts, in all populations and at all time-points, constituting one dataset. Experimental samples are independent both between observation time-points and within each measurement date. Thus, the likelihood function, conditional to the event leading to the infection of $n_{inf}(t)$ hosts at time t is :

$$l(\mathbf{x}|\boldsymbol{\theta}) = \prod_{t \in \mathcal{T}^{obs}} \prod_{p=1}^{n_{inf}(t)} p\left(\mathbf{X}^p(t) = \mathbf{x}^p(t) \mid \boldsymbol{\theta}_{LV}, \boldsymbol{\theta}_{Dir}(t), x_{tot}^p(t)\right).$$

Under our assumption concerning the distribution of the data, we have :

$$p\left(\mathbf{X}^p(t) = \mathbf{x}^p(t) \mid \boldsymbol{\theta}_{LV}, \theta_{Dir}(t), x_{tot}^p(t)\right) = dDM\left(\mathbf{x}^p(t) \mid \boldsymbol{\lambda}^{\boldsymbol{\theta}_{LV}}(t), \theta_{Dir}(t), x_{tot}^p(t)\right),$$

dDM being the probability density function (pdf) of the Dirichlet-multinomial distribution (Kitakado et al., 2006). Under these hypotheses, $\boldsymbol{\theta} = (\boldsymbol{\theta}_{LV}, \boldsymbol{\theta}_{N_e})$ can be inferred after the minimization of $-\log(l(\mathbf{x}|\boldsymbol{\theta}))$, assuming that the mean intrinsic rate of increase \bar{r} of all variants is one, as we are interested only in the relative intrinsic rates of increase. The vector of parameters $\boldsymbol{\theta}_{LV} = \mathbf{r}$ was estimated straightforwardly, setting its lower limit to zero. The values of the vector of parameters $\boldsymbol{\theta}_{N_e}$, of whom dDM depends through $\boldsymbol{\theta}_{Dir}$, were deduced from equation 4.3, defined as the harmonic mean of effective population sizes from inoculation to day t . Optimization was performed on the $\eta(d)$ parameters, representing the mean effective population size from day $\mathbf{T}^{obs}(d-1)$ to day $\mathbf{T}^{obs}(d)$, with $\mathbf{T}^{obs}(0) = 0$. The lower limit of $\eta(d)$ was also set to zero.

The model was implemented in R software version 3.0.2 (R Core Team, 2013), as were all the analyses presented here. The Lotka-Volterra model was simulated with the 'lsoda' function of the 'deSolve' package (Soetaert et al., 2010). Maximum likelihood optimization was carried out with the 'nlminb' optimization routine.

Acknowledgments

We would like to thank Fabio Zanini (Max Planck Institute) and Nicolas Parisey (INRA Rennes) for their invaluable help with FFPopSim and Rcpp, respectively. We also would like to thank Grégory Girardot and Baptiste Lederer for their precious help during experimental manipulations. The simulations were carried out with the Avakas (Bordeaux University) and Migale (INRA Jouy en Josas) computer clusters.

Author Contributions

Conceived and designed the experiments : ER BM AP FF. Performed the experiments : ER BM VS SV FF. Analyzed the data : ER BM LM FG FF. Contributed reagents / materials / analysis tools : ER BM RS AP VS SV FF. Wrote the paper : ER BM LM RS FG FF

Supplementary material

Text S1 : Numerical experiment with FFPopSim

We performed several batches of simulations to assess the ability of the Lotka-Volterra Dirichlet-multinomial model (LVDM) model to infer effective population size and selection coefficients accurately from time-sampled data. We simulated the changes in variant frequencies under the effects of various selection and genetic drift regimes. The simulations were designed to fit our real-life datasets, but the results obtained nevertheless provided some more general insight. We used the forward-time population genetics software FFPopSim to handle these simulations Zanini and Neher (2012a). This software is based on the Wright-Fisher model and is designed for use with haploid individuals and for individuals biallelic at multiple loci. It uses a discrete generation scheme in which mutation, selection and genetic drift are applied to each individual in each generation. It can track the number of individuals carrying each genotype among all possible 2^L variants, where L is the total number of loci. Interestingly, the expected population size can be set at each generation, making it possible to generate any desired time-dependent population size.

The parameters required to run a simulation are the initial variant frequencies ω^{inoc} , the fitness of each of the 2^L variants (vector \mathbf{r}), the mutation rate, and a sequence of instantaneous effective population sizes \mathbf{N} . FFPopSim simulates the vector of variant frequencies at each generation τ , $\omega(\tau)$. Given $\omega(\tau)$, $\omega(\tau + 1)$ is calculated in three steps. The first step is the reweighting of the expected frequencies $\omega(\tau)$ under the effect of selection, according to \mathbf{r} . The expected frequencies are then recalculated to simulate mutations between the 2^L variants, assuming symmetric mutation rates. The outcome of these two steps is the expected frequency $\omega'_i(\tau)$ for each variant i for the next generation. The last step is a resampling of size $N_e(\tau + 1)$ to simulate the effect of genetic drift. For each variant i , the number of individuals participating in the next generation is sampled in a Poisson distribution of mean $N_e(\tau + 1)\omega'_i(\tau)$. Finally, $\omega_i(\tau + 1)$ is assessed as the result of the draw normalized by the sum of all draws for all variants.

For the simulations, we handled three biallelic loci (alleles 0 or 1 at each locus), resulting in $2^3 = 8$ variants. We wished to study $n_{var} = 5$ variants corresponding to variants 100, 010, 001, 110 and 011 (see Fig. S4.8), the infectious clone SON41p being set as 000. We therefore fixed the fitness of the three remaining variants to a value close to 0 and their initial frequency to 0. This ensured that the frequencies of these variants remained negligible, such that these variants were

not considered in the analyses. We then carried out the following three steps : (1) generation of population demogenetic dynamics with given selection and genetic drift parameters, (2) building of numerical datasets with a structure similar to our real-life experiment, by sampling the composition of several populations at various time-points, and (3) estimation of selection and genetic drift parameters with the LVDM model.

Step 1 : Generation of populations demo-genetic dynamics A full factorial design with five factors was used to generate FFPopSim datasets encompassing a wide range of selection and genetic drift regimes. The factor σ_r is the standard deviation of fitness values. Three selection regimes ($\sigma_r = 0.025$ for weak, 0.05 for intermediate and 0.1 for strong selection) were considered, by sampling, for each simulation, the fitness values of $n_{var} - 1$ variants from a normal distribution, $\mathbf{r} \sim \mathcal{N}(\bar{r} = 1, \sigma_r)$, and setting the mean variant frequency \bar{r} to one. Similarly, four factors, N_e^{inoc} , N_e^{IO} , \bar{N}_e^S and σ_S , were used to fine-tune contrasting genetic drift regimes mimicking the three main steps of host infection with a virus. The factor N_e^{inoc} represents the bottleneck at host inoculation, lasting one generation. The factor N_e^{IO} sets the intensity of genetic drift during multiplication in the inoculated organ, over five generations. The factors \bar{N}_e^S and σ_S fine-tune the intensity of genetic drift during the colonization of all organs of the host, this stage being called ‘systemic infection’ and lasting 28 generations. We built the vector \mathbf{N} of single-generation effective population size by drawing independent values of N_e^{inoc} , N_e^{IO} and \bar{N}_e^S from log uniform distributions (table 4.2). More specifically, \bar{N}_e^S is the harmonic mean of systemic effective population sizes. For each dataset, a sequence of seven effective population sizes during systemic infection $N_e^{S_k}$ ($k = 1, \dots, 7$) was drawn from a log-normal distribution, with harmonic mean \bar{N}_e^S and standard deviation σ_S Limbrunner et al. (2000). Each $N_e^{S_k}$ was applied for four generations, considered to be the time required for the colonization of a new organ, leading to a specific effective population size $N_e^{S_k}$ for virus populations. Three values of σ_S were tested, corresponding to weakly to strongly fluctuating effective population size during systemic infection. We thus obtain a vector \mathbf{N} as follows :

$$\mathbf{N} = (N_e^{inoc}, \underbrace{N_e^{IO}, N_e^{IO}, \dots}_{5 \text{ generations}}, \underbrace{N_e^{S_1}, N_e^{S_1}, \dots}_{4 \text{ generations}}, \underbrace{N_e^{S_2}, \dots}_{4 \text{ generations}}, \dots, \underbrace{N_e^{S_k}, \dots}_{4 \text{ generations}}, \dots, \underbrace{N_e^{S_7}, \dots}_{4 \text{ generations}}) \quad (4.6)$$

Overall, we generated 9000 FFPopSim parameter combinations of \mathbf{r} and \mathbf{N} .

The definitions of all factors and associated values are summarized in table 4.2.

Step 2 : Building numerical datasets For each of the 9000 parameter combinations of \mathbf{r} and \mathbf{N} we generated 48 independent FFPopSim simulations corresponding to the dynamics of populations in different hosts. We then carried out virtual observations for eight individual hosts on each of the measurement dates corresponding to those used in the biological experiment, $\mathbf{T}^{obs} = (6, 10, 14, 20, 27, 34)$ dpi. We accounted for the HTS process, by sampling variant frequencies from multinomial distributions of size x_{WF}^{tot} , with frequencies from FFPopSim simulations.

Finally, the dataset generated was accepted according to the following criteria, considered to be satisfied for all real-life experiment datasets (with the exception of one plant in 720). At each measurement date, at least two variants had to be present at a minimum frequency of 1% each in at least 50% of the populations. In addition, at least two variants had to be present in all populations at a minimum frequency of 1% each at the first measurement date (6 dpi). This criterion is hence largely permissive regarding the diversity of virus populations retained (Fig. S4.1).

Step 3 : Estimation of selection and genetic drift parameters with the LVDM model For each accepted dataset, $\boldsymbol{\theta} = (\boldsymbol{\theta}_{LV}, \boldsymbol{\theta}_{N_e})$ was estimated with the LVDM model, using maximum-likelihood methods. We assessed the accuracy of the estimates, by comparing directly the estimated values of intrinsic rates of increase $\hat{\boldsymbol{\theta}}_{LV} = \hat{\mathbf{r}}$ with the true values \mathbf{r} . The $\hat{\boldsymbol{\theta}}_{N_e}$ estimates were compared with the true harmonic mean of effective population sizes, defined at generation τ , as :

$$\bar{N}_e(\tau) = \frac{\tau}{\sum_{k=1}^{\tau} \frac{1}{\mathbf{N}(k)}} \quad (4.7)$$

Bias analyses and goodness-of-fit of the model Relative bias on the estimates of the harmonic mean of effective population sizes was defined as : $B(\bar{N}_e(\tau)) = (\hat{N}_e(\tau) - \bar{N}_e(\tau)) / \bar{N}_e(\tau)$. The overall relative bias on the intrinsic rates of increase \mathbf{r} was evaluated as the mean of variant-specific relative biases : $B(\mathbf{r}) = \sum_{i=1}^{n_{var}} ((\hat{r}_i - r_i) / r_i)$. We also evaluated the goodness-of-fit of the model, by calculating the R squared coefficient of correlation between adjusted and observed (i) mean frequencies, and (ii) standard deviation of frequencies. The adjusted mean frequencies are the $\boldsymbol{\lambda}^{\boldsymbol{\theta}_{LV}}(\tau)$ values from the deterministic Lotka-Volterra model, whereas the adjusted standard deviations of frequencies are given by $\sigma(\boldsymbol{\Lambda}^{\boldsymbol{\theta}_{LV}}(\tau)) = \sqrt{\frac{\boldsymbol{\lambda}^{\boldsymbol{\theta}_{LV}}(\tau)(1 - \boldsymbol{\lambda}^{\boldsymbol{\theta}_{LV}}(\tau))}{1 + \boldsymbol{\theta}_{Dir}(\tau)}}$. The observed means and standard deviations of frequencies were calculated from the data obtained for the eight populations observed per sampling date.

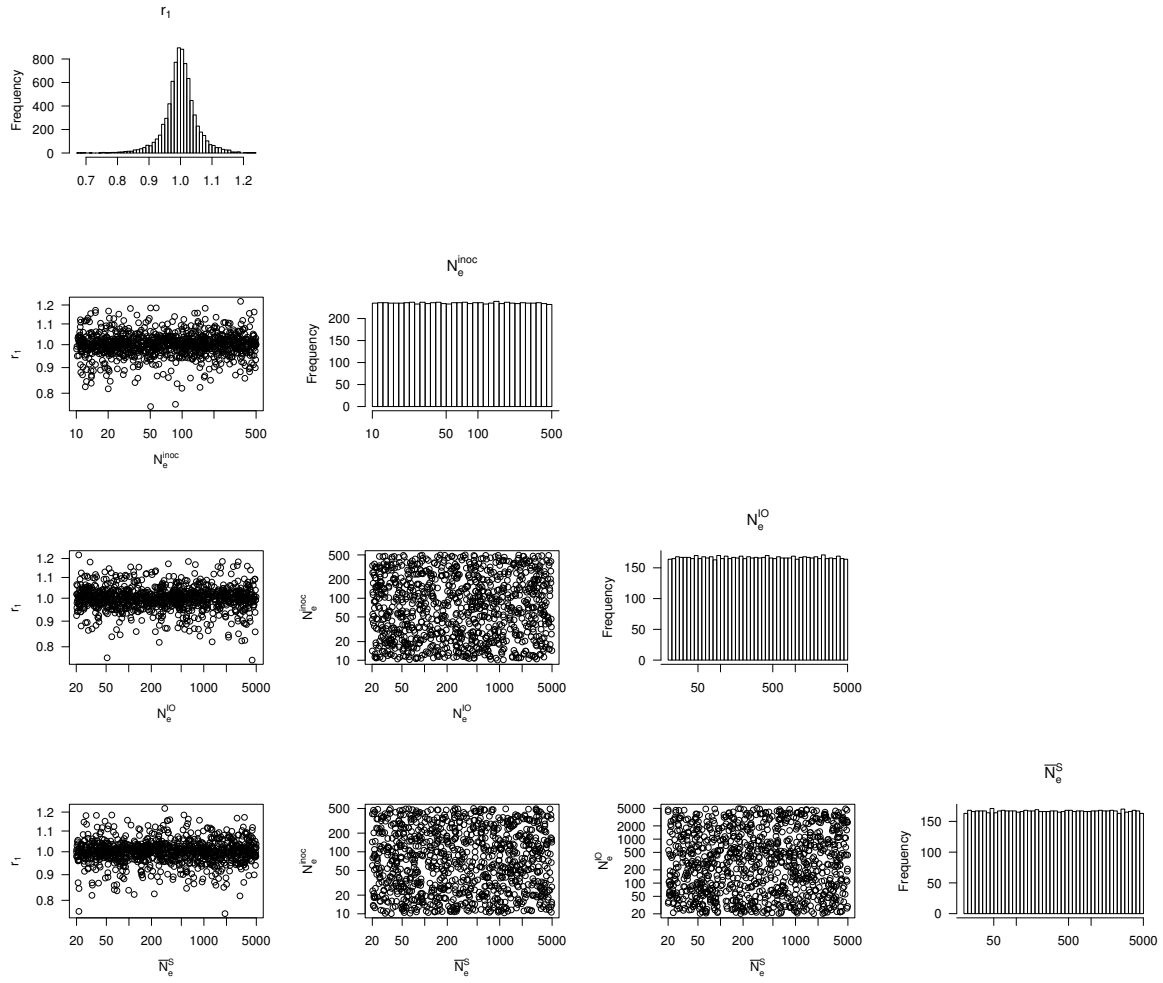


FIGURE S4.1 – **Representation of the variability obtained amongst datasets with FFPopSim.** In the diagonal, each histogram represents the diversity of input parameters r_1 (intrinsic rate of increase of variant 1), N_e^{inoc} (effective population size at inoculation), N_e^{IO} (effective population size in inoculated organ) and N_e^S (harmonic mean of effective population sizes at the systemic level). Under the diagonal, two by two combinations of parameters are represented.

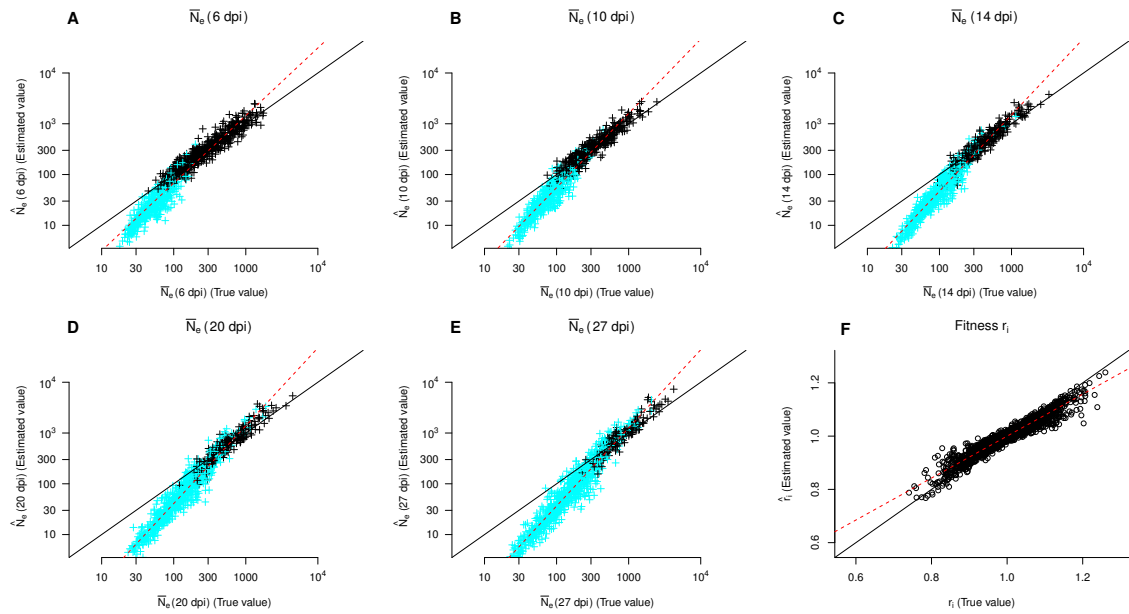


FIGURE S4.2 – Inference of the harmonic mean of effective population sizes \bar{N}_e and variant fitness r with the Lotka-Volterra Dirichlet-multinomial model on datasets simulated with **FFPopSim**. (A-E) Correlation between true \bar{N}_e (x-axis) and predicted \hat{N}_e (raw estimate, y-axis) at 6, 10, 14, 20 and 27 days post-inoculation (dpi). Blue points indicate datasets such that $div(t) < n_{var}$ (with n_{var} the number of variants) and black points indicate datasets such that $div(t) = n_{var}$, $div(t)$ being defined as the mean number of variants with frequencies higher than 1% among the observed populations. (F) Correlation between true r_i (x-axis) and predicted \hat{r}_i (raw estimate, y-axis) (intrinsic rates of increase of variants). In all panels, the black line is the first bisector and the red dashed line is the best linear model fit. For clarity, only 1/3 of the data (randomly drawn) were represented.

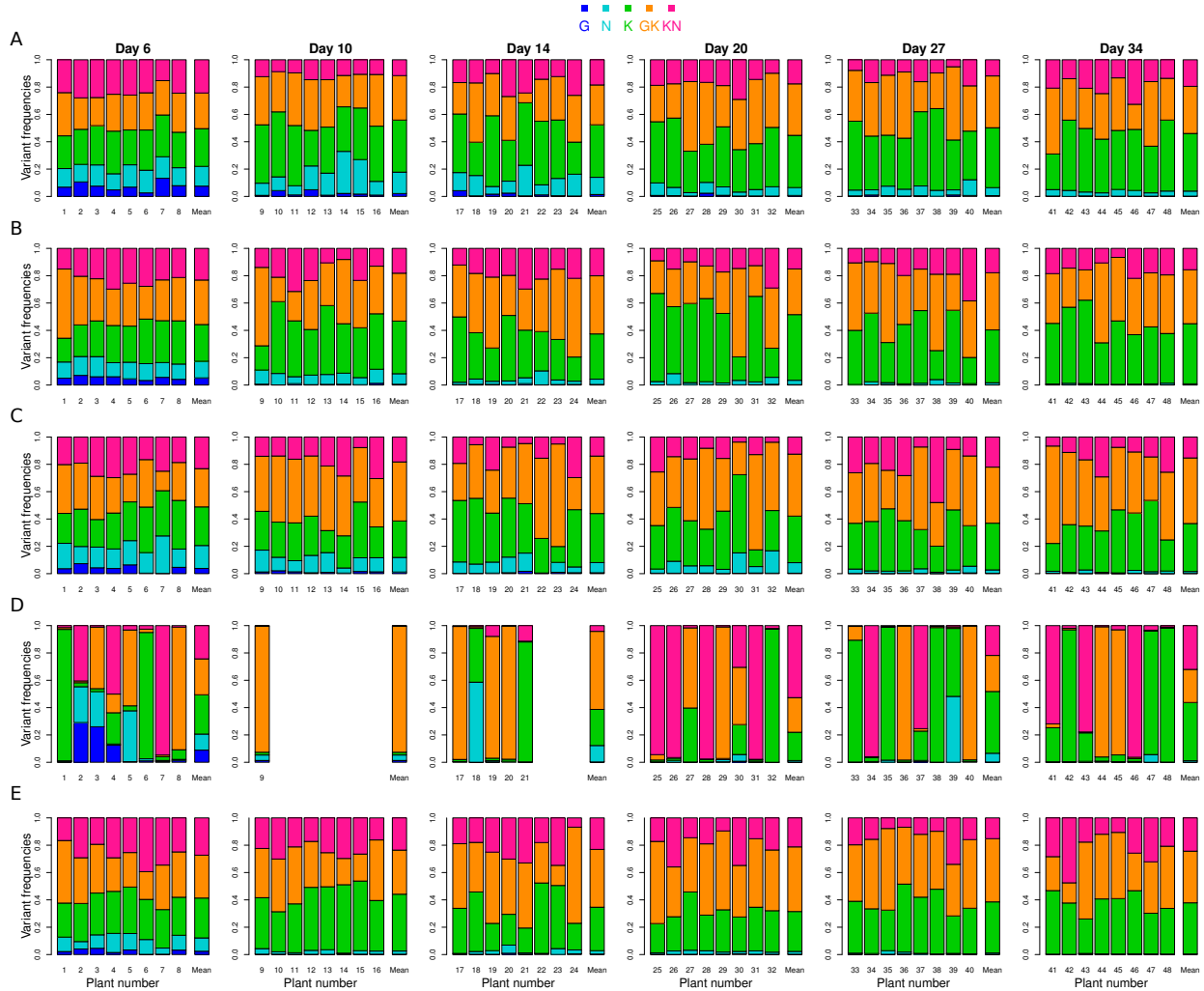


FIGURE S4.3 – **Five datasets obtained by high-throughput sequencing from the biological experiment.** Each line of bar plots represents the dynamics of virus variants in one DH line : (A) 221, (B) 2123, (C) 2173, (D) 2256 and (E) 2264. Within each bar plot, variant frequencies in each plant sample are represented by one bar, except non-infected plants, and the last bar stands for the mean viral composition over the infected plants. One bar plot corresponds to one measurement date, indicated on the top of each column of bar plots. Variant color code is indicated on the top of the figure.

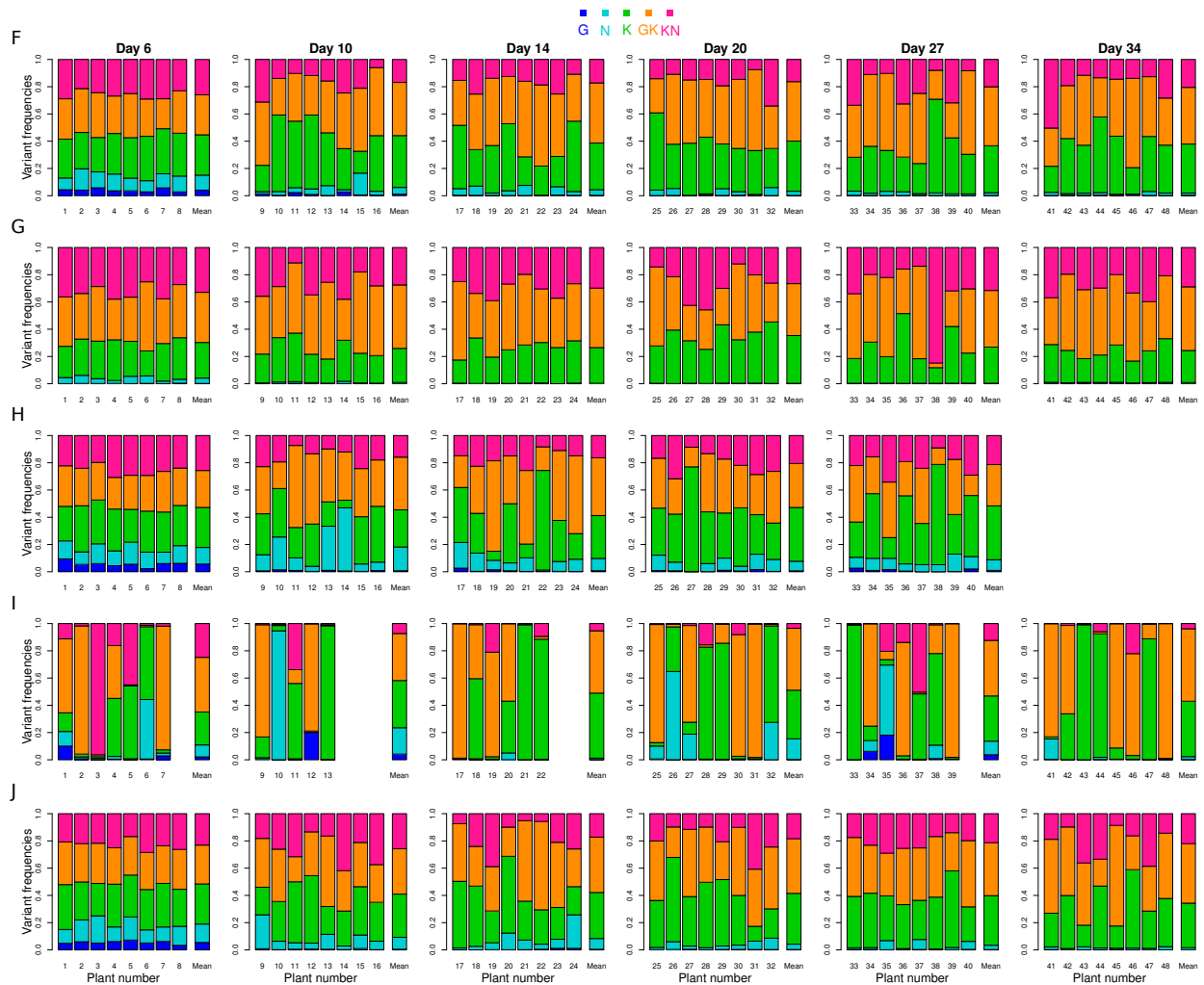


FIGURE S4.4 – Sequel of Fig. S4.3, for DH lines : (F) 2328, (G) 2349, (H) 2367, (I) 2400 and (J) 2426. We could not sample plants of DH line 2367 (H) 34 days post-inoculation because severe necrosis symptoms invading the stem led to all plants death at that late date.

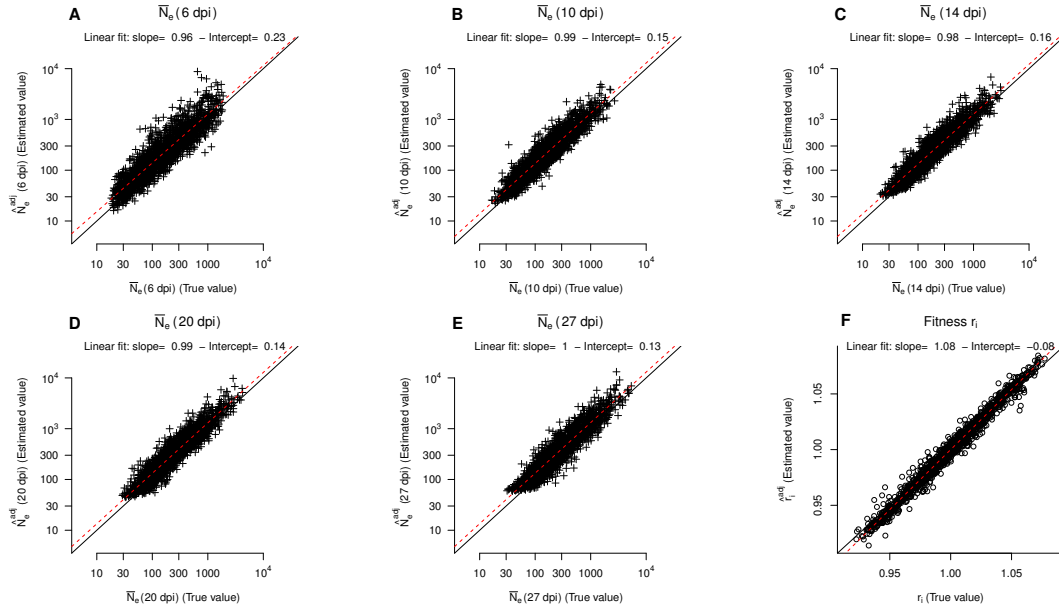


FIGURE S4.5 – Inference of the harmonic mean of effective population sizes \bar{N}_e and variant fitness r with the Lotka-Volterra Dirichlet-multinomial model on the datasets simulated with FFPopSim with two virus variants. These data have been obtained following the three steps detailed in the numerical experiment section except that only two virus variants were considered. (A-E) Correlation between true \bar{N}_e (x-axis) and predicted \hat{N}_e^{adj} (y-axis) at 6, 10, 14, 20 and 27 days post-inoculation (dpi). (F) Correlation between true r_i (x-axis) and predicted \hat{r}_i^{adj} (y-axis) (intrinsic rates of increase of variants). In all the panels, the black line is the first bisector and the red dashed line is the best linear model fit. The slope and the intercept of the best linear fit are indicated in the subtitles. In all, 2000 datasets were generated with FFPopSim in a factorial design similar to the one used for the main numerical experiment with five variants except that only two levels were considered for σ_r (0.025 and 0.1) and one level for σ_S (10^1).

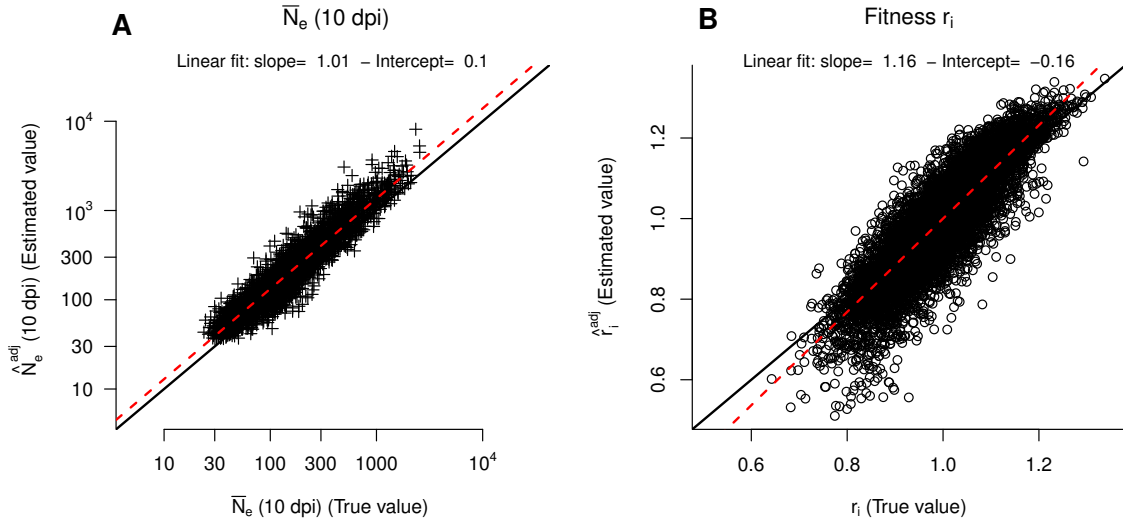


FIGURE S4.6 – **Inference of the harmonic mean of effective population sizes \bar{N}_e and variant fitness r with the Lotka-Volterra Dirichlet-multinomial model on the datasets simulated with FFPopSim with five virus variants and a single sampling date.** These data have been obtained following the three steps detailed in the numerical experiment section except that only 8 plants were observed 10 days post-inoculation (dpi) ($T^{obs} = 10$). (A) Correlation between true \bar{N}_e (x-axis) and predicted \hat{N}_e^{adj} (y-axis) 10 dpi. (B) Correlation between true r_i (x-axis) and predicted \hat{r}_i^{adj} (y-axis) (intrinsic rates of increase of variants). In each panel, the black line is the first bisector and the red dashed line is the best linear model fit. The slope and the intercept of the best linear fit are indicated in the subtitles. In all, 3000 datasets were generated with FFPopSim. The factorial design is similar to the one used for the main numerical experiment except that (i) only one level for σ_S (10^1) was considered and (ii) the criteria for accepting the datasets were more drastic, as follows. Three out of the five variants had to be present in all (of the eight) populations observed at a minimum frequency of 1% each and a fourth variant had to present in 50% of the populations at a minimum frequency of 1%.

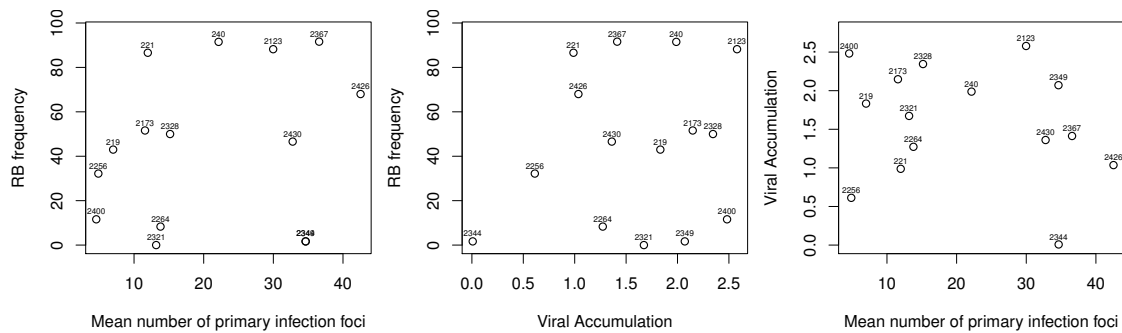


FIGURE S4.7 – **Resistance-breakdown (RB) frequency, viral accumulation and effective population size at inoculation for the 15 DH lines studied.** First panel : RB frequency versus mean number of primary infection foci. Second panel : RB frequency versus viral accumulation. Third panel : viral accumulation versus mean number of primary infection foci. On each graphic, pepper genotypes are represented as points, with their nomenclature (DH line number) given above each point. Zwart et al. (2011) showed that the number of primary infection foci is an accurate estimator of effective population size at inoculation N_e^{inoc} on the pathosystem *Tobacco etch virus* (TEV, genus *Potyvirus*)-*Nicotiana tabacum* plants. We showed that this conclusion is valid for our pathosystem and provided N_e^{inoc} estimates for the 15 DH lines with the *Potato virus Y* (PVY, genus *Potyvirus*) variant K, carrying a green fluorescent marker (green fluorescent protein, GFP). The resistance-breakdown (RB) frequency and the relative viral accumulation were estimated by Quenouille et al. (2013). The RB frequency corresponds to the percentage of infected plants when inoculated with an avirulent variant regarding the allele of resistance *pvr2³*, carried by all DH lines. The relative viral accumulation, or relative viral concentration, was measured by double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA).

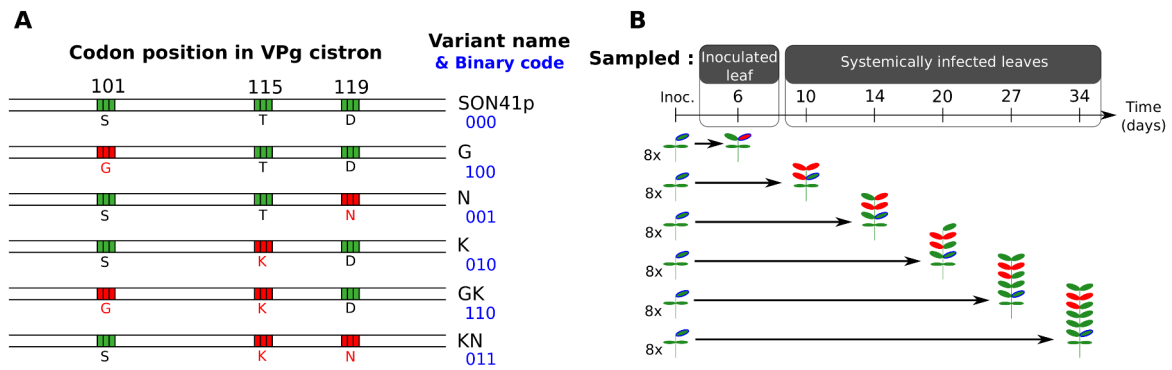


FIGURE S4.8 – **Virus variants inoculated to pepper plants and sampling protocol.** (A) The five virus variants were derived from the SON41p PVY clone and differed only at codon positions 101, 115 and 119 of the VPg cistron. These positions are in green when they correspond to the SON41p clone and in red when a non-synonymous substitution was introduced by site-directed mutagenesis. Single-letter amino-acid abbreviations are presented below each position and PVY variant. Variant names and their respective binary code concerning the three point mutations of interest are given on the right of the sequences, the binary code of the SON41p variant being set to 000. (B) Sampling protocol for one pepper genotype. 48 plants were inoculated, and 8 plants of those 48 were sampled at each measurement time-point, from 6 to 34 days post-inoculation. The blue-circled leaf represents the inoculated leaf. Plain red leaves represent the sampled leaves.

TABLE S4.1 – Tag sequences used to distinguish each plant sample after pooling and MiSeq Illumina high throughput sequencing.

	Name	Polarity	Sequence for binding libraries	Tag sequence	Primer sequence
	Tag 1	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	ACGAGTGCGT	AAGAGAATGTCTATGCTGAC-3'
	Tag 2	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	CATAGTAGTG	AAGAGAATGTCTATGCTGAC-3'
	Tag 3	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	GGTCTAGTAC	AAGAGAATGTCTATGCTGAC-3'
Fwd.	Tag 4	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	TCATACGCGT	AAGAGAATGTCTATGCTGAC-3'
	Tag 5	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	CTACGCTCTA	AAGAGAATGTCTATGCTGAC-3'
	Tag 6	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	ATCGATAGAC	AAGAGAATGTCTATGCTGAC-3'
	Tag 7	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	GAGGCTCTAC	AAGAGAATGTCTATGCTGAC-3'
	Tag 8	+	5'-CTTTCCCTACACGACGCTCTTCCGATCT	TGCTGATATC	AAGAGAATGTCTATGCTGAC-3'
Rev.	REV	-	5'-GGAGTTCAGACGTGTGCTCTTCCGATCT		CAGACCAATCTTTCCTGAAG-3'

The forward (Fwd.) primer sequence was the same for all amplifications and was binded to the sequence tag, just after it. Its binding site corresponds to positions 5971 to 5990 of the PVY isolate SON41p (accession number AJ439544). The binding site of the reverse (Rev.) primer sequence corresponds to positions 6095 to 6114 of the PVY isolate SON41p. RT-PCR amplifications were done according to the following profile : 1h at 42°C, 35 times 10 min at 95°C, 45s at 95°C, 30s at 50°C and 20s at 72°C, and after the cycles 10 min at 72°C.

TABLE S4.2 – Number of sequences and composition of the virus population in each sample of the biological experiment.

In all, 724 samples were analyzed (15 doubled haploid (DH) lines of pepper \times 6 sampling date (dpi : days post-inoculation) \times 8 plants by date and 4 samples for the initial inoculum). Columns indicate (i) the name of each DH line, (ii) the sampling date in dpi, (iii) the number of the sequence tag used (see Table S4.1), (iv) the plant number (as in Fig. 4.3, Fig. S4.3 and S4.4), (v) the infectious status of each sample (0 : not infected / 1 : infected), (vi) the number of cleaned sequences assigned to each sample after filtering and (vii-xi) the number of sequences for each viral variant (G, N, K, GK and KN).

Table in Excel format, available online when the article will be accepted.

Chapitre 5

Impact of genetic drift, selection and accumulation level on virus adaptation to their host plants

List of authors : Elsa Rousseau^{*1,2,3}, Lucie Tami sier^{*3}, Fr d ric Fabre⁴, Vincent Simon^{3,5}, Marion Szadkowski⁶, Gr gory Girardot³, Ludovic Mailleret^{1,2}, Fr d ric Grognard¹, Alain Palloix⁶, Benoit Moury³

¹ Biocore Team, INRIA, F-06902 Sophia Antipolis, France

² INRA, Univ. Nice Sophia Antipolis, CNRS, UMR 1355-7254 Institut Sophia Agrobiotech, 06900 Sophia Antipolis, France

³ UR 407 PV, INRA, F-84140 Montfavet, France

⁴ UMR 1065 SAVE, INRA, F-33882 Villenave d'Ornon, France

⁵ UMR BFP, INRA, F-33882 Villenave d'Ornon, France

⁶ UR GAFL, INRA, F-84140 Montfavet, France

* co-first authors

Abstract

Background

Plant major resistance genes constitute an effective and full protection against viruses, but their efficiency is limited by the emergence and spread of resistance-breaking mutants. Modulating the intensities of evolutionary forces acting on virus populations may constitute a promising tool to increase the durability of major resistance genes. Here, we study the effects of three traits genetically controlled by the host plant on the frequency of resistance breakdown. Two of those three traits, effective population size and selection coefficient, are related to genetic drift and selection evolutionary forces. The third trait, virus accumulation level, can be, to some extent, related to probability of appearance of new mutations. All the variables studied are issued from biological experiments on *Potato virus Y* (PVY) infecting pepper plants, and led to the setting up of two datasets, one on an extended set of pepper genotypes with single time-point estimations and another one on a restricted number of pepper genotypes with an extended number of temporal estimations.

Results

Our study showed that strong genetic drift at early stages of plant infection (i.e. at inoculation and in the inoculated leaf) and low virus accumulation are associated with a decrease in the frequency of resistance-breakdown. No significant correlation was found between selection and resistance-breakdown frequency.

Conclusion

In all, genetic drift is the main factor influencing the frequency of resistance-breakdown. Plant genotypes inducing a small virus effective population size (below 19 per inoculated cotyledon) lead to an increase in resistance durability. Combining these plant genotypes characteristics with ones inducing a small virus accumulation enhance this resistance durability. Breeding plant varieties inducing strong genetic drift and low accumulation on pathogen population hence seems a promising strategy for a more durable management of plant major resistance genes.

5.1 Introduction

Resistance to pathogens, i.e. the capacity of a host to decrease its pathogen load (Restif and Koella, 2004; Råberg et al., 2007), is a widespread defense mechanism in plants. Resistance intensity, specificity and genetic determinism are highly variable. Up to date, plant breeders have mostly created resistant cultivars using resistance mechanisms showing monogenic inheritance. These monogenic resistances are often called qualitative resistance as they are controlled by a major resistance gene displaying a high efficiency. Unfortunately, the protection against pathogens conferred by such resistance genes was often poorly durable (García-Arenal and McDonald, 2003; McDonald and Linde, 2002a). Indeed, in most of the reported cases, ten cropping seasons with plant cultivars carrying a given resistance gene are sufficient for the counter-adaptation of the targeted pathogen, impairing resistance efficiency (García-Arenal and McDonald, 2003). Usually, this resistance breakdown involves the selection by the plants of pathogen variants carrying mutation(s) in one, or a small number, of specific pathogen genes.

Three main evolutionary processes drive pathogen evolution : mutation, selection and genetic drift. Mutation creates new variants in a pathogen population, adapted to a hitherto efficient major resistance gene (McDonald and Linde, 2002a). Selection favors the pathogen variants with highest fitness, increasing their frequencies from one generation to the next, at the expense of the weakest ones. This deterministic force is usually evaluated with the selection coefficient s , defined as the difference in fitness between two alleles or two variants. By contrast, genetic drift acts in the same way on all variants of a pathogen population, introducing random fluctuations in the dynamics of variant frequencies (Charlesworth, 2009). This stochastic force is commonly evaluated with the effective population size (N_e), defined as the size of an idealized population (i.e. a panmictic population of constant size with discrete generations) that would show the same degree of randomness in the evolution of variant frequencies as the population under consideration (Kimura and Crow, 1963; Wright, 1931).

The breakdown of a major resistance gene recently introgressed into commercial varieties and used by growers can be schematically divided into three major steps that must be successfully completed by the pathogen (Moury et al., 2011b). If we assume that no resistance-breaking pathogen variant is present initially, the first step consists in the appearance of resistance-breaking variants from a wild-type pathogen population. As soon as these resistance-breaking variants appear, they are in competition with the other components of the pathogen population for the colonization of the plants where they first appeared. Depending on the intensities of genetic drift and selection, the resistance-breaking variants will accumulate enough and colonize

these plants (step 2). Then, they may have the opportunity to be transmitted to other plants and to spread in crops (step 3), allowing epidemics to develop in plants carrying the resistance. Different evolutionary forces rule these three steps. In the case of plant viruses, appearance of resistance-breaking variants usually involves a small number of nucleotide substitutions in the so-called avirulence factor encoded by the viral genome (Harrison, 2002; Moury et al., 2011b). Exceptionally, recombination may be required for appearance of resistance-breaking virus variants (Díaz et al., 2004; Miras et al., 2014). Then, accumulation of the resistance-breaking variants within plants depends on its relative fitness, i.e. on the selection coefficient, and on the intensity of genetic drift modulated by the bottlenecks that occurs at multiple steps of plant infection (Gutiérrez et al., 2010, 2012a; Zwart and Elena, 2015). Modeling approaches have forecasted that the evolutionary forces acting at the within-plant scale, especially the mutational pathway involved in resistance breakdown and the fitness cost associated with the resistance-breaking mutation(s) accounted for about 50% of the risk of resistance breakdown in the field (Fabre et al., 2009, 2012b, 2015). Experimental data have also shown that these two factors were good predictors of the risk of resistance breakdown (Fabre et al., 2012a; Harrison, 2002; Janzac et al., 2009). The remaining 50% depended on factors related to epidemiology and thus to step 3 of resistance breakdown.

One way to avoid or delay the breakdown of monogenic qualitative resistances is to combine the major resistance gene with a suitable plant genetic background. Plant accessions carrying both a major resistance gene and a partially-resistant genetic background showed a highly significant increase in major resistance gene durability compared to accessions carrying only the major resistance gene, in the case of plant interactions with a RNA virus (Palloix et al., 2009), a fungus (Brun et al., 2010) or a nematode (Fournet et al., 2013). Indeed, the genetic background of the host can be composed of quantitative trait loci (QTLs) that affect the level of resistance to pathogens and the intensity of evolutionary forces acting on pathogen populations (Lannou, 2012). When combined with a major resistance gene, the genetic background of a host plant may have three main effects, detailed thereafter from studies of Quenouille et al. (2013, 2014, 2016).

In the case of the *Potato virus Y* (PVY, genus *Potyvirus*) - pepper (*Capsicum annuum*; family Solanaceae) interaction, Quenouille et al. (2013, 2016) showed a significant correlation between the breakdown frequency of a major resistance gene and the capacity of the virus to accumulate in the plant, i.e. the additional resistance level conferred by the plant genetic background. They hypothesized that within-plant virus accumulation was linked to the overall number of virus genome replications during plant infection as well as to the population growth

rate, and consequently to the probability of appearance of the resistance-breaking mutations in the virus population, assuming that the intrinsic mutation rate per replication was identical between plant genotypes. Using a progeny of 153 pepper genotypes carrying the same major resistance gene but contrasted genetic backgrounds, Quenouille et al. (2014) mapped QTLs controlling virus accumulation within-plants or the frequency of breakdown of the major resistance gene in the pepper genome. This mapping revealed that the two QTLs controlling virus accumulation colocalized with QTLs controlling the frequency of breakdown of the major resistance gene, which provided a genetic explanation for the observed correlation between the two traits.

By comparing two pepper genotypes carrying the same major resistance gene associated with either a partially-resistant genetic background or a susceptible background, Quenouille et al. (2013) also showed that the selection of the adapted resistance-breaking PVY mutants was slower in the former accession. The slower selection may be a consequence of the smaller selection coefficient of the adapted mutants, of a more intense genetic drift in the plants with a partially-resistant genetic background, or of both (Quenouille et al., 2013). Consequently, in the analyzed PVY - pepper system, the mutation, selection and/or genetic drift evolutionary forces acting on the virus population could determine the breakdown or durability of the major resistance gene.

To sum up, the three main hypothetical effects of the plant genetic background on the durability of a major resistance gene identified by Quenouille et al. (2013, 2016) are : (i) a decrease in the probability of appearance of the mutation(s) involved in resistance breakdown in the pathogen population, because of a decrease in the pathogen census population size, (ii) a slower selection of the pathogen mutants carrying the resistance-breaking mutations once they have appeared, because of a reduced selection coefficient between pathogens carrying or not resistance-breaking mutations, or because of a reduced pathogen effective population size, counteracting the effect of selection, and (iii) an increased extinction probability of the resistance-breaking mutants through more intense genetic drift during plant infection, because of small pathogen effective population size.

The aim of the present study was to determine in the same PVY - pepper system which of the three factors, virus accumulation within-plants, selection coefficient between virus variants and virus effective population size, was (or were) determining the breakdown of the major resistance gene and to estimate their relative importance in order to help the breeding of future plant cultivars with durable pathogen resistance.

5.2 Results

5.2.1 Estimation of the frequency of breakdown of the major resistance gene *pvr2³* by PVY and of putative explanatory variables linked to PVY evolution

The resistance breakdown frequency at the individual plant level (RB) and several putative explanatory variables linked to the mutation, selection and genetic drift evolutionary processes were estimated in 151 doubled haploid (DH) pepper lines (*Capsicum annuum*) – or in subsets of these 151 DH lines – all carrying the *pvr2³* major resistance allele but differing in their genetic background. Different variants of the wild-type PVY isolate SON41p (Moury et al., 2004) were chosen for such estimations, depending on the explanatory variables. The explanatory variables, detailed thereafter, concerned PVY accumulation within-plants, selection coefficients between variants and PVY effective population sizes at several stages on the infection. We investigated the links between these evolutionary variables and RB .

RB was previously evaluated after inoculation of each of the 151 DH lines (60 plants per DH line) with the 'CI chimera', an artificial recombinant of SON41p carrying the cylindrical inclusion (CI)-coding region of PVY isolate LYE84.2 (experiment 1, Table 5.1) (Montarry et al., 2011; Quenouille et al., 2014). This variant was preferred to SON41p because of its higher ability to break the *pvr2³* resistance, providing a larger range of RB values among DH lines, hence allowing a higher accuracy for genetic and statistical analyses. Still, RB obtained with SON41p was shown to be highly correlated with RB obtained with the CI chimera on a subset of sixteen contrasted DH lines (Quenouille et al., 2013).

RB corresponds to the frequency of plants showing virus infection at the systemic level around one month after inoculation and was shown to correspond to situations where a nonsynonymous mutation in the VPg-(viral protein genome-linked) coding region became fixed in the PVY population, conferring adaptation to the *pvr2³* resistance (Ayme et al., 2006; Montarry et al., 2011). By contrast, a PVY variant carrying none of these nonsynonymous mutations will be qualified as wild-type, like for SON41p or the CI chimera. The terminology 'resistance breakdown' is frequently used to describe the increase of infection rates, and often of subsequent economic losses, in crops of plant cultivars carrying resistance genes following the adaptation of pathogen populations. For simplicity, we use this term to describe pathogen adaptation at the individual plant level, without any assumption about the consequences at the field scale or in terms of losses. In this experiment, a high heritability ($h^2 = 0.87$) was shown for RB

(Quenouille et al., 2014).

Effective population size at plant inoculation (N_e^{In}) was estimated on the 151 DH lines with a SON41p variant carrying a GFP (Green Fluorescent Protein) reporter gene and a single amino acid substitution in the VPg (threonine to lysine substitution at codon position 115; 'K' substitution) which allows infection of plants carrying *pvr2³* (experiment 2, Table 5.1) (Tamisier et al., 2016). N_e^{In} corresponds to the mean number of PVY primary infection foci visualized by the GFP fluorescence 5 or 6 days after mechanical inoculation of 20 pepper cotyledons per DH line. N_e^{In} was also shown to have a high heritability ($h^2 = 0.93$).

TABLE 5.1 – Description of the variables used in this study.

Variable ^a	Dataset ^b	Exp. ^c	Infection stage	Measurement date	Number of DH lines	Inoculum ^d	Reference
RB	1 & 2	1	Systemic	38 dpi ^e	151	CI chimera	Quenouille et al. (2014)
N_e^{In}	1 & 2	2	Inoculation	5 & 6 dpi	151	SON41p-K : :GFP	Tamisier et al. (2016)
VA^S	2	3	Systemic	36 dpi	151	CI chimera-N	Quenouille et al. (2014)
N_{e1}^{IL}	1	4	Inoculated leaf	6 dpi	15	SON41p-G, N, K, GK, KN	Rousseau et al. (Submitted)
N_{e1}^S	1	4	Systemic	20 dpi	15	SON41p-G, N, K, GK, KN	Rousseau et al. (Submitted)
s_1	1	4	All	6 to 34 dpi	15	SON41p-G, N, K, GK, KN	Rousseau et al. (Submitted)
VA_1^{IL}	1	4	Inoculated leaf	6 dpi	15	SON41p-G, N, K, GK, KN	Present study
VA_1^S	1	4	Systemic	10 to 27 dpi	15	SON41p-G, N, K, GK, KN	Present study
N_{e2}^S	2	5	Systemic	21 dpi	90	SON41p-G, N, K, GK, KN	Present study
s_2	2	5	All	21 dpi	90	SON41p-G, N, K, GK, KN	Present study

^a Figures 1 and 2 were used as indices for variables estimated in experiments number 4 and 5, respectively.

^b Two datasets corresponding to different numbers of DH lines were analyzed separately in the article.

^c Different figures correspond to independent experiments. Variables sharing the same number have been estimated on the same plants.

^d One or two letter codes correspond to amino acid substitutions in PVY VPg allowing infection of plants carrying the *pvr2³* resistance gene.

^e dpi : days post-inoculation.

Relative PVY accumulation in plants was measured in infected leaves by quantitative DAS-ELISA (double antibody sandwich enzyme-linked immunosorbent assay) (Quenouille et al., 2014). In the 151 DH lines, mean relative virus accumulation (VA) was assessed in pools of three systemically-infected leaves per plant at 36 days post-inoculation (dpi), with one experimental replica comprising 10 plants per DH line (variable VA^S , experiment 3, Table 5.1) (Quenouille et al., 2014). For this, a mutant of the CI chimera of SON41p carrying the aspartic acid to asparagine substitution at amino acid position 119 of the VPg ('N' substitution) that allowed infection of plants carrying $pvr2^3$ was used. A fairly good heritability was observed for VA^S ($h^2 = 0.64$) (Quenouille et al., 2014).

Based on the three variables RB , N_e^{In} and VA^S , 15 contrasted DH lines (Fig. S5.1) were chosen for an accurate estimation of the intensity of genetic drift and selection (experiment 4, Table 5.1). The experiment consisted in following the dynamics of an initially nearly equimolar population of five SON41p VPg mutants with one or two nonsynonymous substitutions (named G, K, N, GK and KN, Rousseau et al. (Submitted)) in the mechanically-inoculated leaf at 6 dpi, and in pools of three systemically-infected leaves at 10, 14, 20, 27 and 34 dpi, using MiSeq Illumina high-throughput sequencing (HTS). Again, each of these mutations or mutation pairs (double-letter names) allowed the infection of plants carrying $pvr2^3$ ('K' and 'N' substitutions defined previously; 'G' substitution : serine to glycine substitution at codon position 101 of the VPg). The PVY effective population size at each sampling time-point and the selection coefficients between the five PVY mutants were accurately estimated with a new mechanistic-statistical model (Rousseau et al., Submitted). The relative intrinsic rate of increase of each PVY mutant was estimated based on the evolution of its mean frequency over eight plants at each time-point. The selection coefficient in each DH line, an explanatory variable denoted s_1 , was set as the difference in relative intrinsic rates of increase between the fittest and the weakest variants. The dynamics of effective population sizes N_e were estimated from the variance of mutant frequencies between plants at each time-point. Thereafter, we will mainly focus on N_e in the inoculated leaf at 6 dpi, denoted N_{e1}^{IL} , and N_e in systemically-infected leaves at 20 dpi, denoted N_{e1}^S (Table 5.1).

The model was validated on numerical data generated from a Wright-Fisher model (Zanini and Neher, 2012a,b) with known genetic drift and selection parameters in Rousseau et al. (Submitted). This validation showed that model predictions are highly accurate both for effective population sizes ($R^2 \in [0.93, 0.95]$) and selection coefficients ($R^2 = 0.94$). Additionally, the heritability of the traits, calculated on the fifteen DH lines, was highly satisfactory ($h^2 = 0.94$

for effective population size and $h^2 = 0.79$ for selection coefficients).

From aliquots of the same plant samples, the relative PVY accumulation, VA, was measured by quantitative DAS-ELISA at the same time-points, i.e. at 6, 10, 14, 20, 27 and 34 dpi. Thereafter, we will mainly focus on VA in the inoculated leaf at 6 dpi, denoted VA_1^{IL} , and on VA in systemically-infected leaves from 10 to 27 dpi, VA_1^S , calculated as the non-standardized area under the virus accumulation curve between these time-points.

A fifth experiment was conducted to jointly estimate effective population size and selection coefficient on a larger host population of 90 DH lines issued from the same progeny as the 15 DH lines described before (Table 5.1). Due to the much higher number of plant genotypes than in the previous experiment and in order to fit with the size of the Illumina MiSeq run, sampling of plants was performed at a single date (21 dpi) in all systemically-infected leaves on eight plants per genotype. Initially, the whole set of 151 DH lines were inoculated, but no-infection cases and amplification and contamination issues left us with 90 DH lines (see Materials and Methods). The same model as the one used with experiment 4 was employed to estimate the relative fitness of virus variants and the strength of genetic drift (Rousseau et al., Submitted). The validity of model estimations with only one sampling date (plus inoculum) was also tested with numerical simulations (Rousseau et al., Submitted). Predictions were satisfactory both for the effective population size ($R^2 = 0.89$) and the selection coefficient ($R^2 = 0.80$). Thereafter, we will denote the effective population size in systemically-infected leaves at 21 dpi N_{e2}^S and the selection coefficient between the fittest and the weakest variants for each DH line s_2 .

Even if experiments 4 and 5 are overlapping at some points, they both have their advantages and drawbacks, which makes them both useful and complementary for our purposes. On the one hand, experiment 4 on a reduced number of DH lines provides more accurate estimations, and captures the time dimension of the infection process, distinguishing inoculation and/or inoculated leaf colonization from systemic infection. Indeed, we could estimate dynamics of N_e and of VA, while we had to focus on one time-point for experiment 5, chosen as a late systemic infection date. On the other hand, experiment 4 has a low number of plant genotypes (15), whereas experiment 5 dealt with a much higher number of plant genotypes (90), providing more power in the upcoming statistical analyses.

Two datasets were constituted based on the DH lines of experiments 4 and 5. Dataset 1 is based on the 15 DH lines of experiment 4 and dataset 2 on the 90 DH lines of experiment 5. Details about the variable of each dataset are provided in Table 5.1.

TABLE 5.2 – Correlations between explanatory variables and the resistance-breakdown frequency, for dataset 1 including 15 DH lines.

	VA_1^{IL}	N_{e1}^{IL}	VA_1^S	N_{e1}^S	N_e^{In}	s_1	RB
VA_1^{IL}	-	0	0.70	$2.2 \cdot 10^{-3}$	$6.9 \cdot 10^{-3}$	$2.0 \cdot 10^{-2}$	$1.4 \cdot 10^{-3}$
N_{e1}^{IL}	0.87*	-	0.43	0	$3.6 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$2.3 \cdot 10^{-3}$
VA_1^S	0.11	0.22	-	0.11	0.19	0.98	0.24
N_{e1}^S	0.74*	0.91*	0.43	-	0.17	$7.7 \cdot 10^{-3}$	$1.4 \cdot 10^{-2}$
N_e^{In}	0.68*	0.55*	-0.36	0.37	-	$1.4 \cdot 10^{-2}$	0.42
s_1	0.60*	0.65*	0.01	0.67*	0.63*	-	0.54
RB	0.75*	0.73*	0.32	0.62*	0.23	0.17	-

Variables are described in Table 5.1. Correlation coefficients $\rho_{Spearman}$ are below the diagonal and p_{values} corresponding to the null hypothesis $\rho_{Spearman} = 0$ are above the diagonal. Correlations corresponding to variable RB are shown in the grey column and line. * next to $\rho_{Spearman}$ values indicate significant Spearman correlations ($p_{value} \leq 0.05$).

5.2.2 Correlation between putative explanatory variables

Table 5.2 shows the correlations among the explanatory variables estimated with dataset 1 on 15 DH lines, VA_1^{IL} , N_{e1}^{IL} , VA_1^S , N_{e1}^S , N_e^{In} and s_1 . Those 15 pepper genotypes were chosen for showing contrasting values of N_e^{In} , VA^S and RB , with the objective to observe a large panel of genetic drift and selection patterns. It is important to note that, as a consequence of this choice, the correlations between the three variables N_e^{In} , VA^S and RB were not significant for that reduced dataset (Table 5.2, Fig. S5.1). The same explanation holds also certainly for the lack of correlation between VA_1^S and RB , or between VA_1^S and N_e^{In} (Table 5.2). Because of this, it is not relevant to compare those variables in dataset 1 to understand their biological relationships.

Regarding effective population sizes, we observed that the correlation between N_e at two successive infection stages is significant but it is not when infection stages are more distant. Indeed, N_e measured at inoculation is significantly correlated with N_e measured in the inoculated leaf ($\rho_{Spearman} = 0.55$) and N_e measured at the systemic level is strongly correlated with N_e measured in the inoculated leaf ($\rho_{Spearman} = 0.91$). However, the correlation is not significant when comparing the inoculation stage with the systemic infection stage.

The effective population size is significantly correlated with VA when both are measured at the same date from 6 to 14 dpi (Fig. 5.1), but the correlation is lost when comparing N_e and VA at 20 or 27 dpi (Table S5.1, Spearman $p_{value} \geq 0.35$) or when comparing the effective population size at 20 dpi (N_{e1}^S) and VA_1^S measured from 10 to 27 dpi (Table 5.2). Also, the correlation is strong between VA in the inoculated leaf and effective population size in apical leaves, but the opposite is not true. Coming back to the time-point estimations of N_e and VA at 6, 10, 14, 20 and 27 dpi, the link between VA at one date to N_e at the next date is in

fact significant until N_e estimated at 20 dpi, whereas it is until 14 dpi between VA and N_e at the same date (Table S5.1). At the opposite, the link between N_e at one date and VA at the next date is poor as it is significant only for N_e estimated at 6 dpi. Interestingly, the effective population size at inoculation is significantly correlated with VA in the inoculated leaf, but not with VA in apical leaves.

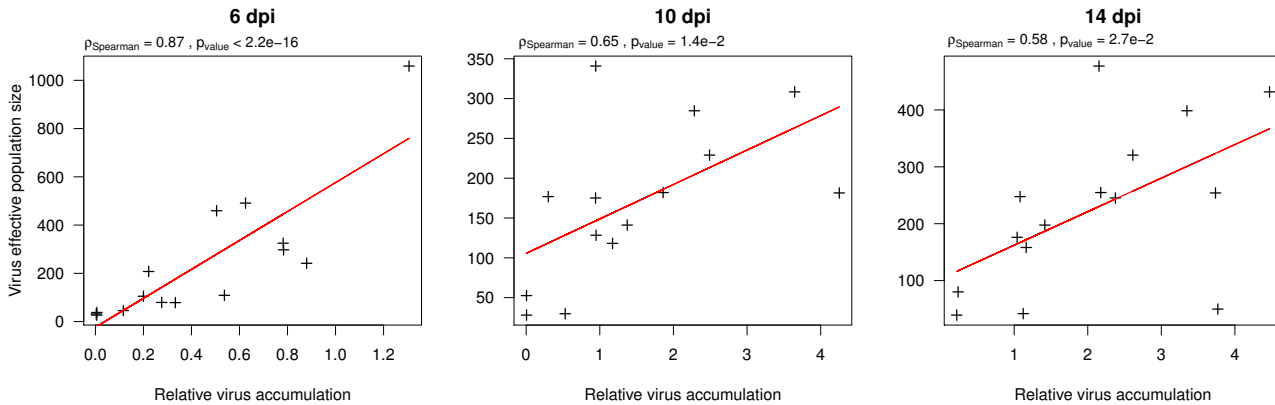


FIGURE 5.1 – **Link between PVY effective population size and relative virus accumulation at 6, 10 and 14 days post-inoculation (dpi)**. Correlation between virus effective population size, inferred from the model in Rousseau et al. (Submitted), and the relative viral concentrations at 6, 10 and 14 dpi, for the 15 DH lines of experiment 4 (see Table 5.1). The red line is the best fit line (linear model fit). The p_{value} and ρ_{Spearman} estimate of a correlation test are indicated on top of each graphic.

The three PVY effective population sizes, i.e. at inoculation, in the inoculated leaf and during systemic infection, are significantly correlated with s_1 . From these correlations, we could identify two groups of DH lines : one group with a low selection coefficient and a small effective population size, and one group with a high selection coefficient and a large effective population size (data not shown).

The selection coefficient is correlated with VA in the inoculated leaf, even though to a limited extent, but no correlation is observed between s_1 and VA during systemic infection.

The correlations between all explanatory variables associated with dataset 2 (VA^S , $N_e In$, s_2 and N_{e2}^S) are reported in Table 5.3. As for dataset 1, the effective population size at inoculation is not correlated with the one during systemic infection. VA^S is correlated with the effective population size in apical leaves, but not with the one at inoculation. Similarly, the selection coefficient is not correlated with the effective population size at inoculation, but it is with the one during systemic infection. However, the distribution of the points in the space (N_{e2}^S , s_2) for the 90 DH lines of dataset 2 (Fig. 5.2) was slightly different from what was observed on the reduced set of DH lines in dataset 1 (between N_{e1}^S and s_1). Indeed, in dataset 1 the positive

TABLE 5.3 – Correlations between explanatory variables and the resistance-breakdown frequency, for the DH lines dataset 2.

	VA^S	N_e^{In}	s_2	N_{e2}^S	RB
VA^S	-	0.15 ‡	0.42	$3.6 \cdot 10^{-2}$	$3.6 \cdot 10^{-7}$ ‡
N_e^{In}	0.12	-	0.52	0.24	$2.3 \cdot 10^{-3}$ ‡
s_2	0.09	0.07	-	$2.2 \cdot 10^{-3}$	0.60
N_{e2}^S	0.22*	0.12	0.32*	-	0.71
RB	0.40*	0.25*	-0.06	0.04	-

Variables are described in Table 5.1. Correlation coefficients $\rho_{Spearman}$ are below the diagonal and p_{values} corresponding to the null hypothesis $\rho_{Spearman} = 0$ are above the diagonal. Correlations corresponding to variable RB are shown in the grey column and line.

* Significant Spearman correlations ($p_{value} \leq 0.05$).

‡ Correlation test on 151 DH lines (90 DH lines otherwise; Table 5.1), indicated only next to p_{values} for simplicity.

correlation between the effective population size and the selection coefficient allowed (N_e , s) estimations to vary between small N_e and s to large N_e and s , whereas on the larger set of DH lines of dataset 2, some DH lines with both a small effective population size during systemic infection and a large selection coefficient were observed. Finally, s_1 is not correlated with VA^S .

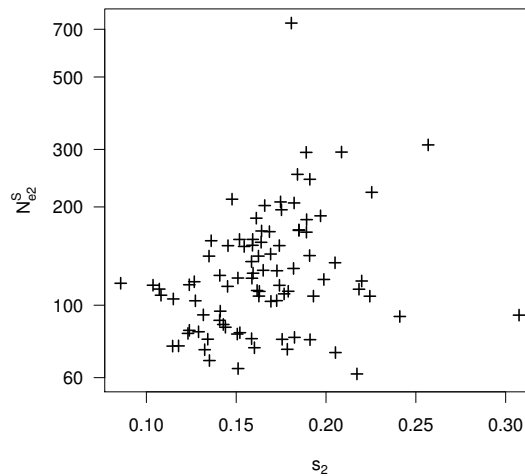


FIGURE 5.2 – Link between the effective population size in systemically-infected leaves N_{e2}^S and selection coefficient s_2 . Correlation between PVY effective population size in systemically-infected leaves at 21 days post-inoculation N_{e2}^S and the selection coefficient s_2 inferred from the model by Rousseau et al. (Submitted), for the 90 DH lines of experiment 5 (dataset 2). Correlation coefficient $\rho_{Spearman}$ is equal to 0.32 and the corresponding p_{value} is equal to $2.2 \cdot 10^{-3}$.

Finally, all significant correlations are positive (Tables 5.2 and 5.3). Hence, for those significant cases, an increase in N_e , VA or s is associated with an increase in another one of those three variables, or the same variable at another stage of the infection.

5.2.3 Analysis of the links between the frequency of breakdown of the major resistance gene and putative explanatory variables

We analyzed the correlations between RB and the potential explanatory variables linked to the selection and genetic drift forces, and to the census population size. For dataset 1, one can observe that RB is significantly correlated with the effective population size, both in the inoculated leaf and to a lower extent during systemic infection (Table 5.2). Regarding VA , the correlation with RB is strong in the inoculated leaf at 6 dpi, but not significant for the systemic infection stage. Finally, no correlation between RB and the s_1 is noticed. For dataset 2, the correlation between RB and N_e^{In} , but also with VA^S , are significant. On the contrary, correlation is not significant between RB and N_e^S or s_2 . For both datasets, all significant correlations between RB and N_e , VA or s are positive, indicating that an increase in one of those explanatory variables is associated with an increase in RB .

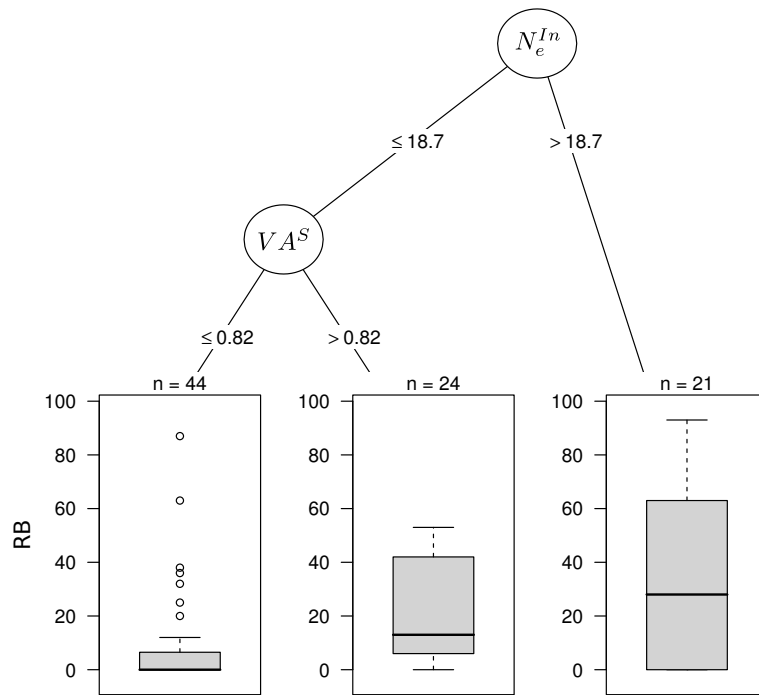


FIGURE 5.3 – **Regression tree on resistance breakdown frequencies for dataset 2 including 90 DH lines.** Explanatory variables were : virus accumulation at 36 days post-inoculation (dpi) VA^S , effective population size at inoculation N_e^{In} , selection coefficient s_2 and effective population size at 21 dpi N_e^S . Only two explanatory variables (VA^S and N_e^{In}) out of four ended in the regression tree. Boxplots of RB are represented for the n observations in each terminal node.

Given the larger number of observations in dataset 2, we could compute a regression tree to explain RB with all four explanatory variables alone (VA^S , N_e^{In} , s_2 , N_e^S). Even though no constraint was applied to compute the regression tree, only two variables were kept, N_e^{In}

and VA^S . The first-level variable in the tree was N_e^{In} . If N_e^{In} is large (> 18.7), then RB is predicted to be high (median at around 30% and first and third quartile confidence interval at approximately $[0, 60]\%$). If N_e^{In} is small (≤ 18.7), we get to a node corresponding to the second-level variable, VA^S . If VA^S is large (> 0.82), then in average RB is predicted to be intermediate (median at around 15%, first and third quartile confidence interval at approximately $[5, 40]\%$). Finally, if VA^S is small (≤ 0.82), then RB is predicted to be low (median at around 0%, first and third quartile confidence interval at approximately $[0, 5]\%$).

5.3 Discussion

5.3.1 Genetic drift and resistance efficiency are complementary levers to improve the durability of virus resistance genes

The main objective of this article was to study the relationships between the frequency of breakdown of the major resistance gene $pvr2^3$ and proxy of the main underlying evolutionary forces of genetic drift and selection, and of the census population size. Indeed, we assumed that some of these evolutionary forces may be used to predict the frequency of breakdown of the major resistance gene (Quenouille et al., 2013). In the two datasets analyzed, the VA level and the intensity of genetic drift, especially at the early infection steps (i.e. at inoculation or in the inoculated leaf), were significantly correlated with the frequency of resistance breakdown (Tables 5.2 and 5.3). On the contrary, no significant correlation with resistance breakdown was observed for the selection coefficient and a weaker (dataset 1) or a lack of (dataset 2) correlation was observed for the intensity of genetic drift at the systemic level.

The importance of VA on resistance breakdown may be the consequence of its link with the probability of appearance of the resistance-breaking mutations in the inoculated plants (Quenouille et al., 2013). The underlying hypothesis of this statement is that the within-plant accumulation of the resistance-breaking variants – that was used to measure VA_1^{LL} , VA_1^S and VA^S (Table 5.1) – is correlated to the residual accumulation of the wild-type variant in these plant genotypes. We could not measure the accumulation of the wild-type variant because resistance-breaking variants emerged rapidly in most of these plants (Montarry et al., 2011). However, previous observations suggested that the major resistance gene ($pvr2^3$) present in the analyzed plants does not control completely PVY multiplication and even movement within the plant (Montarry et al., 2011). In 1% of PVY-inoculated plants that carried the $pvr2^3$ resistance and that get infected at the systemic level was the sequence of the PVY population

identical to that of the wild-type (i.e. non-resistance-breaking) inoculum. In the majority of the infected plants, these wild-type viruses were outcompeted by the resistance-breaking mutants and could not be observed. In addition, Montarry et al. (2011) argued that the resistance-breaking PVY mutants identified in infected plants had appeared in the inoculated plants and did not preexist in the inoculum, and the frequency of breakdown is therefore linked to the residual accumulation of the wild-type PVY variant in these plants. Altogether, these results suggest that the resistance-breaking mutants appear in the inoculated plants, probably most frequently in the inoculated organs.

Under this scenario, the observed positive correlation between resistance breakdown and VA or N_e at the early stages of infection (Tables 5.2 and 5.3) may be a joint effect of (i) delaying the virus infection and/or decreasing VA, and (ii) purging the resistance-breaking mutants from the virus population of the inoculum, consequently reducing the probability of appearance and/or spread of the resistance-breaking mutants.

Regarding the lack of correlation between s and RB, this result may be attributable to our measure of s . We could not evaluate the selection coefficient between the wild-type variant and a resistance-breaking variant in plants carrying the major resistance gene because the tremendous competition would rapidly lead to the extinction of the wild-type variant (Montarry et al., 2011). As a consequence, we calculated s as the difference in relative fitness between the fittest and the weakest resistance-breaking variants, which may be not informative for the risk of resistance-breaking. Previous studies on the same pathosystem consisted in estimating the fitness of PVY variants (wild-type clone SON41p and resistance-breaking variants) in competition experiments between SON41p and a resistance-breaking variant, or in single infection experiments (Ayme et al., 2006, 2007). The authors showed that fitness estimations from single infection experiments fitted better their experimental data of resistance breaking than fitness estimations from competition experiments.

5.3.2 Complex interplays between virus within-host accumulation, genetic drift and selection

Within plants, there are complex interplays between mutation, genetic drift and selection processes acting on virus populations. Genetic drift is caused by the numerous narrow bottlenecks operating at the virus inoculation step, or during within-cell multiplication and cell-to-cell or systemic movements (Gutiérrez et al., 2010, 2012a; Zwart and Elena, 2015). Genetic drift can slow down adaptation (Charlesworth, 2009; Kimura, 1970; Rouzine et al., 2001) or even decrease

the overall fitness of a population because of the accumulation of deleterious mutations, a process known as Muller's ratchet (Chao, 1990; de la Iglesia and Elena, 2007; Duarte et al., 1992; Lázaro et al., 2003). However, under some circumstances, genetic drift can also favor adaptation, as suggested in the case of a synergistic interaction for fitness between genotypes in a virus population (Miyashita and Kishino, 2010) or in the case of complex adaptive trajectories when the fitness landscape is rugged (Wright, 1932; Zwart and Elena, 2015). Indeed, using a model of competition among intracellular virus populations, Miyashita and Kishino (2010) showed that narrow bottlenecks during cell-to-cell movement help isolating adaptive genomes from defective ones, allowing selection to operate efficiently. Zwart and Elena (2015) argued that, when the fitness of the two single-mutants is lower or higher than the fitness of the wild-type variant and the double-mutant (reciprocal-sign epistasis) (de Silva and Wyatt, 2014; Lalić and Elena, 2012), which is characteristic of rugged fitness landscapes with multiple fitness peaks (Poelwijk et al., 2011), narrow bottlenecks during virus infection can provide some advantages for their adaptation. This follows from the fact that successive bottlenecks can allow the fixation of slightly deleterious mutations, and thus potentially allow the virus to cross a fitness valley and reach a higher fitness peak. Nevertheless, our results do not support the hypotheses of Miyashita and Kishino (2010) and Zwart and Elena (2015) of a beneficial effect of genetic drift for adaptation, but rather a beneficial effect of genetic drift to slow down adaptation. More generally, there are complex relationships between the number of mutations required for adaptation, the probability of appearance of each of these mutations, the incurred fitness changes and genetic drift (de Silva and Wyatt, 2014; Fabre et al., 2009; Iwasa et al., 2004; Quenouille et al., 2013). This emphasizes the importance of accurately estimating selection and genetic drift acting on virus populations and their census population sizes in order to predict the risk of resistance-breaking virus variant emergence.

The probability of appearance of a given mutation in a plant virus depends on its intrinsic mutation rate, for which only few estimates have been experimentally obtained (de la Iglesia et al., 2012; Gago et al., 2009; Malpica et al., 2002; Tromas and Elena, 2010), on the type of mutation (Ayme et al., 2006; Quenouille et al., 2013), on the population growth rate and on the generation time, which is poorly known (Khelifa et al., 2010; Martínez et al., 2011). Because of the experimental difficulty of estimating the probability for a new mutation to occur in a given number of generations, we used the VA level as a proxy. Indeed, the VA level is an indicator of the census population size at a given time-point, which depends on the population growth rate and generation time. Assuming that the intrinsic mutation rate is similar between DH lines,

the VA level would actually be a proxy of the probability for a new mutation to occur in a given number of generations.

Several of the explanatory variables we studied, i.e. N_e , VA and s , were correlated (Tables 5.2 and 5.3). First, we observed significant positive correlations between N_e and s intensities within-hosts (Table 5.2). The correlation was particularly strong ($\rho_{Spearman} \geq 0.63$) for dataset 1 and was moderate ($\rho_{Spearman} = 0.32$) for dataset 2. This suggests a trade-off between genetic drift and selection, explained in population genetics by the fact that a small effective population size reduces the genetic variability in a population and the effectiveness of selection (Charlesworth, 2009). However, several DH lines showed both high s and low N_e values (Fig. 5.2), demonstrating that the genetic control of N_e and s intensities by the host plant (Rousseau et al., Submitted) can drive these intensities beyond this trade-off. Also, Lanfear et al. (2014) argued that a small effective population size decreases the rate of appearance and fixation of adaptive mutations.

Second, there was a significant correlation between N_e at inoculation and VA in the inoculated leaf (Table 5.2) but not with VA at the systemic level (Tables 5.2 and 5.3). Using the potyviruses *Tobacco etch virus* (TEV) or *Turnip mosaic virus* (TuMV) and *Nicotiana* spp. plants, a low N_e at inoculation was shown to result in a delay in plant infection at the systemic level (Rodrigo et al., 2014), or in a lower proportion of infected cells (Lafforgue et al., 2012), hence probably also in a lower VA. Our results show the same tendency but only between inoculation and inoculated leaf colonization. The discrepancy could be due to the plant genotypes used for the experiments, either highly susceptible to virus infection (Lafforgue et al., 2012; Rodrigo et al., 2014) or partially resistant (this study). Partial resistance mechanisms acting specifically on the virus systemic movement may be the cause of the lack of relationship between the density of infection foci in the inoculated leaf and the efficiency of systemic infection. This is consistent with the observation of a narrow bottleneck at 10 dpi, i.e. at the onset of systemic infection (Rousseau et al., Submitted), whereas such a bottleneck did not seem to occur in the experiments performed with TEV and *Nicotiana tabacum* (Zwart et al., 2011).

Another interesting observation is that N_e measured at one date is best correlated with VA measured at the previous date than VA measured at the same date or at the next date (Table S5.1). This observation posits an alternative view from the results of Rodrigo et al. (2014) and Lafforgue et al. (2012) previously mentioned. Indeed, this suggests a causal relationship between VA at one date and N_e at the next date and not the other way around. We can explain

this with the compelling example of the multiple bottlenecks experienced by viruses during plant infection (Gutiérrez et al., 2010, 2012a; Zwart and Elena, 2015). Bottlenecks are events corresponding to a reduction of the census population size, hence also of VA. It is clear that a bottleneck (reduction in VA) will lead to an increase in genetic drift, hence to a reduction in N_e , and not the other way around. Also, when the population increases in size (increasing VA), genetic drift becomes weaker and N_e becomes larger. Supporting this point of view, Zwart et al. (2011) found that the TEV effective population size after the inoculation bottleneck depends on the dose of the inoculum.

The fact that the correlation between VA and N_e decreases with time and becomes non-significant after 14 dpi (Fig. 5.1) in dataset 1 may be due to the fact that we sampled the same amount of infected leaves (3 leaves) at all dates while the plants were bearing an increasing amount of leaves as they were growing. As a consequence, the virus population sampled became less and less representative of the whole population with time, decreasing the accuracy of VA and N_e estimates. Also, as N_e corresponds in fact to estimations of harmonic means of effective population sizes from inoculation to the date of interest (Rousseau et al., Submitted), the link between N_e and VA is expected to blur with time notably because the value of N_e is strongly influenced by previous strong genetic drift periods affecting virus populations.

We emphasize the consistency between the results obtained with the two datasets, that possess opposite advantages and drawbacks in terms of accuracy and statistical power regarding the correlations observed between the explanatory variables (Tables 5.2 and 5.3).

5.3.3 Concluding remarks

Overall, our results suggest that the intensity of genetic drift is an effective trait to promote the durability of plant major resistance genes to viruses (Zhan et al., 2015), as proposed more generally to limit the emergence of microbe variants adapted to drug treatments (Abel et al., 2015). In our case, this is particularly true for genetic drift acting in the early steps of infection. In addition, an increased intensity of genetic drift would benefit from being combined with a decrease of within-plant accumulation, to minimize the risk of resistance breakdown at the within-plant scale. The combination of these two factors, i.e. VA and N_e in the inoculated leaf for dataset 1 and VA at the systemic level and N_e at inoculation for dataset 2, and their pairwise interaction explains 69% and 20% of the variability of resistance breakdown, respectively, all factors being highly significant. The regression tree based on dataset 2 (Fig. 5.3) provides a practical framework to optimize the breeding for enhanced resistance durability.

N_e at inoculation is the prime factor : When it exceeds a given threshold (here, about 19 infection foci per inoculated cotyledon), resistance breakdown is quite frequent on average. In contrast, when N_e at inoculation is below this threshold, resistance durability is higher and can still be enhanced when combined with a reduced VA.

5.4 Materials and methods

5.4.1 Analysis of composite PVY populations infecting pepper DH lines

Experiment 5 included, among others, the same doubled haploid (DH) lines of *C. annuum* and was performed on the same five PVY variants as those used by Rousseau et al. (Submitted). We used 151 DH lines carrying the PVY resistance allele *pvr2³* and differing in their genetic background. The five PVY variants, named G, N, K, GK and KN based on their amino acid substitutions in the VPg, were mixed in equimolar amounts based on quantitative DAS-ELISA and mechanically inoculated to the two cotyledons of 8 plants per DH line. The major difference from the experiment by Rousseau et al. (Submitted) was that sampling was performed at a single date, at 21 days post-inoculation, and all apical leaves were collected. For each plant, all sampled leaves were pooled and crushed in buffer before RNA purification, as in Rousseau et al. (Submitted). One-step reverse-transcription polymerase chain reaction (RT-PCR) amplification was conducted for the 1208 plants (8 plants \times 151 DH lines) individually, in thirteen 96-well PCR plates. In all, the sampling, RT-PCR and sequencing procedures performed to obtain the viral sequences were essentially as described by Rousseau et al. (Submitted).

Three of the eight primers used by Rousseau et al. (Submitted) were poorly efficient in PCR (primer tag 5'-GGTCTAGTAC, tag 5'-GAGGCTCTAC and tag 5'-TGCTGATATC), and were thus replaced with primers 5'-CGACGACTGC, 5'-TGGAGTACGA and 5'-GGAGCGTCAC, respectively. Single-step RT-PCR was performed on RNA extracts on a first set of 13 PCR plates. Additionally, for 66 DH lines, no RT-PCR products were detected by agarose gel electrophoresis for at least 4 of the 8 plants. A two-step RT-PCR protocol was thus carried out for all the samples of these 66 DH lines in six 96-well PCR plates. This protocol usually provides a higher sensitivity than the first one but increases the risk of contamination (Bustin, 2000). Our experiment was in agreement with Bustin (2000)'s report, since we observed contaminations in the negative controls for 4 of the 6 additional PCR plates.

As next-generation sequencing data may be impacted by contaminations and potential

biases during PCR amplification, several controls were conducted. Potential cross-contamination among samples during the crushing step in a roller mill was evaluated thanks to plants of the pepper genotype Yolo Wonder, mechanically inoculated with the SON41p G' mutant (aspartic acid to glycine substitution at codon position 119 of the VPg) which differs by at least two amino acids from the five variants studied. Every 56 leaf samples, one leaf from these infected plants and put through the roller mill. The samples crushed right after these leaves were used as contamination controls and tested for the presence of G' variant sequences. Twenty-three contamination controls were performed in total. Illumina MiSeq sequencing revealed that no sequences belonging to the G' variant were detected in all contamination control, indicating that no cross-contamination occurred among the samples during the crushing step.

Additionally, to control for contaminations during RT-PCR, 8 negative controls (i.e. one per primer pair), where the RNA template was replaced with water, were added on each PCR plate. After Illumina MiSeq sequencing, the number of reads detected in the negative controls of the former set of thirteen PCR plates ranged from 0 to 100, with a mean number (\pm standard deviation) of 40 ± 24 . The sequencing results confirmed the occurrence of high contamination levels on the four plates of the latter set of six PCR plates, with 1449 ± 1917 reads in the negative controls. Consequently, all DH lines corresponding to these four PCR plates were removed from the final dataset. Finally, 90 DH lines were kept. The number of reads per sample of these DH lines ranged from 201 to 11052, with a mean number of 4919 ± 1747 . The initial inoculum was also sequenced and the following frequencies were obtained for the five PVY variants : G (22%), N (15%), K (17%), GK (18%) and KN (28%).

5.4.2 Quantitative estimation of PVY accumulation in pepper plants

An experiment by Rousseau et al. (Submitted) has consisted in following the genetic dynamics of an artificial composite population comprising the five PVY mutants G, K, N GK and KN in 15 DH lines, by sampling virus populations and analyzing their composition in the inoculated leaf at 6 dpi and in three apical leaves at 10, 14, 20, 27 and 34 dpi. Eight plants were sampled per DH line at each time-point, and the frequencies of the PVY mutants were assessed by HTS. An aliquot of each crushed plant sample was collected and the eight aliquots corresponding to the same host genotype and the same measurement date were pooled together. Relative VA were then measured by quantitative DAS-ELISA, as described in Quenouille et al. (2014). VA_1^S was set as the non-standardized area under the VA progress curve from 10 to 27 dpi (Table 5.1). We did not include N_e and VA at 34 dpi in the analyses presented in this

article because very strong symptoms prevented us from carrying out measurements at that date in some DH lines.

5.4.3 Inference of virus effective population size and selection coefficients

Estimation of the strength of genetic drift and selection acting on virus populations was performed on PVY mutant frequencies obtained from HTS (experiments 4 and 5, Table 5.1). For this purpose, we used the Lotka-Volterra Dirichlet-multinomial (LVDM) model taking its roots in Fabre et al. (2012a) and further developed by Rousseau et al. (Submitted). It is composed of (i) deterministic Lotka-Volterra equations of competition between virus mutants, allowing to infer the intrinsic rate of increase r_i of each mutant i , and (ii) a stochastic Dirichlet-multinomial process, allowing to infer variance effective population size $N_e(t)$ at each time-point t . For a complete description of the model and its equations, see Rousseau et al. (Submitted). The selection coefficient between the two mutants i and j corresponds to the difference in their intrinsic rates of increase. In this study, S_1 and s_2 were calculated as the difference between the highest and the lowest intrinsic rates of increase among the five variants for each DH line.

5.4.4 Statistical analyses of the links between variables related to the evolution of PVY populations

All analyses were handled with the R software version 3.0.2 (R Core Team, 2013). An exploratory analysis was conducted, consisting in Spearman's correlation tests between RB (response variable) and putative explanatory variables and also among explanatory variables. Additionally, a regression tree was realized based on dataset 2 to better understand the effect of the four explanatory variables VA^S , N_e^{In} , s_2 and N_{e2}^S alone on RB . It was performed thanks to the 'party' package, without any restriction on the maximum depth of the tree or on the minimum number of observations in terminal nodes.

Acknowledgements

The authors would like to thank AM Sage-Palloix and G Nemouchi for providing pepper genetic resources, and N Truglio's team for technical assistance.

Supplementary Material

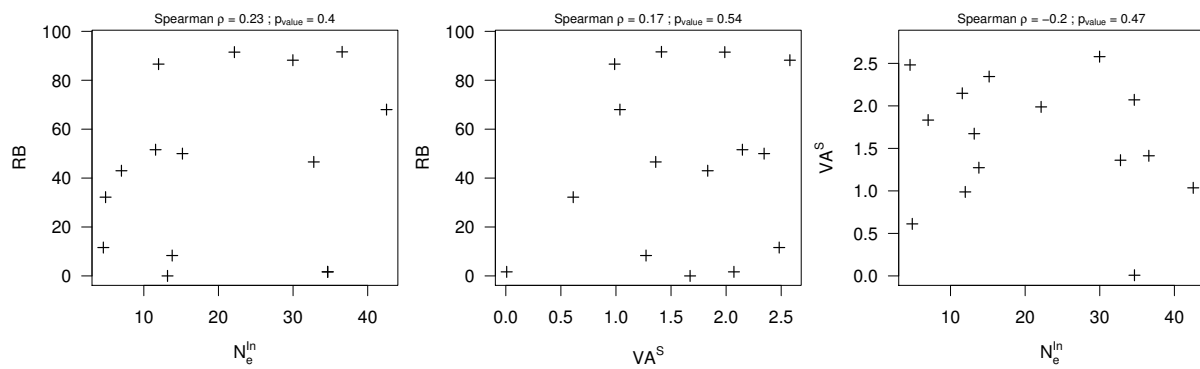


FIGURE S5.1 – Characteristics of the 15 pepper genotypes used in the experiment 4 (Table 5.1) in terms of effective population size at inoculation (N_e^{In}), resistance breakdown frequency (RB) and virus accumulation during systemic infection (VA^S) (modified from Rousseau et al. (Submitted)). On each graphic, each pepper genotypes is represented by one cross. Note that the three variables are not significantly correlated.

TABLE S5.1 – Correlation between experiment 4 estimations (Table 5.1) of the effective population size N_e and virus accumulation VA at 6, 10, 14, 20 and 27 days post inoculation (dpi), at one sampling date difference at the most.

	$N_e(6 \text{ dpi})$	$N_e(10 \text{ dpi})$	$N_e(14 \text{ dpi})$	$N_e(20 \text{ dpi})$	$N_e(27 \text{ dpi})$
$VA(6 \text{ dpi})$	0.87*	0.70*			
$VA(10 \text{ dpi})$	0.59*	0.65*	0.70*		
$VA(14 \text{ dpi})$		0.48	0.58*	0.58*	
$VA(20 \text{ dpi})$			0.26	0.26	0.22
$VA(27 \text{ dpi})$				0.11	0.086

* indicate significant Spearman correlations ($p_{value} \leq 0.05$).

Chapitre 6

Synthèse, discussion et perspectives

6.1 Synthèse des principaux résultats

Ce manuscrit de thèse repose principalement sur l'étude de la dérive génétique chez les virus de plantes, et comment cette dérive génétique peut permettre d'augmenter la durabilité d'un gène majeur de résistance. Ces travaux font suite aux hypothèses évoquées par Quenouille et al. (2013) concernant la plus grande durabilité du gène majeur de résistance *pvr2³* lorsqu'il est en présence d'un fond génétique conférant un niveau de résistance quantitative plutôt qu'en présence d'un fond génétique sensible (Palloix et al., 2009) (Fig. 1.14). En suivant la piste des nombreux goulets d'étranglement identifiés chez les virus de plantes, depuis la transmission inter-hôtes jusqu'à la colonisation d'un hôte (Gutiérrez et al., 2012a; Zwart and Elena, 2015), j'ai voulu étudier l'effet de la dérive génétique sur le risque d'émergence de variants virulents face à une résistance qualitative (García-Arenal and McDonald, 2003; McDonald and Linde, 2002a). J'ai mené plusieurs études, depuis le niveau intra-hôte, faisant appel à des expériences biologiques couplées à leur modélisation ou à des analyses statistiques, jusqu'au niveau d'une parcelle agricole, par une approche théorique des processus épidémiologiques.

Un modèle épidémiologique stochastique au niveau d'une parcelle agricole a permis de prédire l'effet de la dérive génétique sur la durabilité d'un gène majeur de résistance et sur les gains de rendement additionnels pouvant être obtenus. L'utilisation d'un modèle stochastique plutôt que déterministe est justifiée par la présence de charges virales faibles aux goulets d'étranglement. Dans de tels cas, les prédictions de modèles stochastique et déterministe peuvent différer, et un modèle stochastique représente par définition mieux les fluctuations aléatoires correspondantes (Bertacchi et al., 2015). La dérive génétique a été introduite dans le modèle en s'inspirant des multiples travaux ayant montré l'existence de nombreux goulets d'étranglement

subis par les populations virales depuis la transmission inter-hôtes jusqu'au début de l'infection systémique d'un hôte (Gutiérrez et al., 2012a; Zwart and Elena, 2015). Ce modèle simule le déploiement d'une variété de plante pyramidant un gène majeur de résistance avec une résistance quantitative diminuant les tailles des goulets d'étranglement subis par les populations virales. D'après ce modèle, l'effet moyen de la dérive génétique est soit bénéfique, soit neutre, voire marginalement négatif pour les gains de rendement agricoles. La dérive génétique peut être particulièrement bénéfique lorsque la (ou les) mutation nécessaire à l'adaptation du virus au gène majeur induit un coût de fitness intermédiaire dans les plantes sensibles. Lorsque le coût de fitness est faible, cela se traduit par une fréquence élevée des variants virulents dans les plantes sensibles à l'équilibre sélection-mutation, et donc une probabilité élevée que ces variants soient transmis aux plantes résistantes, même en réduisant les tailles de goulets d'étranglement. À l'opposé, lorsque le coût de fitness est fort, les variants virulents sont en fréquence très faible dans les plantes sensibles, ce qui se traduit par une probabilité très faible qu'ils soient transmis aux plantes résistantes, même lorsque les goulets d'étranglement sont larges. Ainsi, pour des coûts de fitness intermédiaires, le gène majeur de résistance seul n'est ni peu, ni très efficace, et donc la dérive génétique peut jouer un rôle bénéfique pour les gains de rendement et pour retarder le contournement du gène majeur.

Dans l'étude correspondant au chapitre 4, j'ai amélioré et validé le modèle mécanistico-statistique de Fabre et al. (2012a), permettant d'estimer conjointement la dynamique des tailles efficaces de population $N_e(t)$ et les coefficients de sélection s entre variants viraux, à partir de données temporelles de compétition entre variants viraux dans des plantes. La validation du modèle s'est basée sur des simulations numériques du modèle de Wright-Fisher avec N_e et s connus, générant ainsi des ensembles de données de dynamique de variants viraux en compétition, suivies de l'estimation avec notre modèle des paramètres sur ces jeux de données simulés. Les estimateurs débiaisés de N_e et s développés se sont montrés très satisfaisants, et le modèle s'est révélé robuste pour différents nombres de dates d'échantillonnage (1 ou 6), et différents nombres de variants viraux en compétition (2 ou 5). Ainsi, des estimations fiables de N_e et s ont pu être obtenues à partir de données expérimentales issues du séquençage haut-débit de dynamiques de populations virales constituées de cinq variants viraux en compétition dans différentes lignées HD de piment. Grâce à ces estimations, j'ai pu montrer une forte variabilité dans les intensités de dérive génétique et de sélection agissant sur les populations virales d'une lignée HD à l'autre, et que ces deux caractères étaient contrôlés génétiquement par les plantes. Les lignées HD étudiées étant toutes porteuses du même gène majeur de résistance $pvr2^3$ mais différant par leur fond génétique, le contrôle de N_e et s provenait donc du fond génétique.

Ainsi, il est possible de jouer sur l'intensité de la dérive génétique et de la sélection, par un choix judicieux du génotype de plante.

L'étape suivante a consisté à se demander s'il existe un lien expérimental entre l'intensité de la dérive génétique et/ou de la sélection et la durabilité du gène majeur *pvr2³*, comme proposé par Quenouille et al. (2013) (Fig. 1.14). Cette question est l'objet du chapitre 5 de ce manuscrit. Pour y répondre, j'ai effectué des analyses de la corrélation entre les estimations de N_e et s du chapitre 4, ainsi que d'autres estimations sur un plus grand nombre de lignées HD mais effectuées sur une seule date d'échantillonnage, et les estimations de la fréquence de contournement du gène majeur *pvr2³* obtenues par Quenouille et al. (2014). J'ai également inclus, parmi les variables pouvant expliquer de la fréquence de contournement, les estimations d'accumulation virale obtenues à partir des échantillonnages correspondant à l'expérience du chapitre 4, et d'autres estimations à une date unique tardive mais effectuées sur un plus grand nombre de lignées HD, obtenues par Quenouille et al. (2014) (Table 5.1). Globalement, des corrélations positives et significatives ont été mises en évidence entre N_e aux dates précoces de l'infection (inoculation ou feuille inoculée) et la fréquence de contournement de *pvr2³*, ainsi qu'entre l'accumulation virale et la fréquence de contournement. Cela signifie que des génotypes de plante diminuant N_e ou l'accumulation virale, proxy de la taille réelle de population, permettraient de diminuer la fréquence de contournement du gène majeur *pvr2³*, et donc d'augmenter sa durabilité. Comme supposé par Quenouille et al. (2013), une diminution de la taille réelle de population, estimée par l'accumulation virale, s'accompagne d'une diminution de la probabilité d'apparition des mutations responsables du contournement d'un gène majeur, ce qui expliquerait son lien avec la fréquence de contournement (Fig. 1.14). Aussi, confirmant une hypothèse de Quenouille et al. (2013) (Fig. 1.14), j'ai trouvé une corrélation positive et significative entre la taille réelle de population virale et N_e . À l'inverse, aucun lien significatif n'a été identifié entre s et la fréquence de contournement. Comme évoqué dans le chapitre 5, ce résultat est peut-être imputable à notre mesure de s , que nous avons calculé comme la différence de fitness entre deux variants virulents, plus ou moins bien adaptés dans des plantes portant le gène majeur *pvr2³*. Il est probable que dans chaque plante un seul mutant soit apparu et ait envahi la population virale dans l'expérience de Quenouille et al. (2013), alors que nous avons mis cinq variants en compétitions. Aussi, les mutations ayant conduit au contournement ont pu être diverses d'une plante à l'autre dans l'expérience de Quenouille et al. (2013), et potentiellement différentes de celles que nous avons étudiées. Ayme et al. (2006, 2007) ont constaté que la fitness de variants viraux infectant seuls les plantes était mieux corrélée à la fréquence de contournement que la fitness estimée lorsque les variants viraux étaient en compétition. On pourrait donc étudier

directement la fitness de chaque variant viral infectant indépendamment chaque plante plutôt que des différences de fitness entre variants viraux en compétition, comme dans le cas du paramètre s . Le modèle d'estimation de N_e et s proposé dans le chapitre 4 requiert la présence de plusieurs variants viraux en compétition, mais la fitness d'un variant infectant seul une plante peut être approximée simplement par son accumulation virale.

Au final, j'ai pu mettre en évidence le potentiel effet bénéfique de la dérive génétique pour de meilleurs gains de rendement dans les cultures, conférés par l'utilisation de variétés porteuses d'un gène majeur de résistance. Les résultats prometteurs du modèle épidémiologique sont renforcés par le lien fort identifié entre les données expérimentales de tailles efficaces de population virale à des dates précoces de l'infection et les fréquences de contournement du gène majeur *pvr2³*. L'approche étudiée pour agir sur l'intensité de la dérive génétique est originale puisqu'elle se base sur l'utilisation de résistances quantitatives diminuant les tailles des goulets d'étranglement subis par les populations virales. En effet, la forte variabilité identifiée entre lignées HD pour N_e pourrait être exploitée dans les programmes d'élaboration de variétés, pour sélectionner des génotypes conduisant à davantage de dérive génétique. Une telle stratégie associée à un déploiement adéquat pourrait permettre une meilleure gestion de la résistance des plantes, ainsi que des gains de rendement plus élevés.

6.2 Discussion

Au cours de cette partie, je vais discuter des aspects de ma thèse que j'estime être les plus importants, notamment des résultats ou des constructions de modèles et d'expériences. Je mettrai cette discussion en relation avec quelques articles que j'ai sélectionnés pour, selon mon point de vue, leur pertinence par rapport à ma thèse et les questions qu'ils amènent.

6.2.1 Intensité de la dérive génétique et taille de la population d'hôtes

Bergstrom et al. (1999) rapportent que, pour un agent pathogène transmis horizontalement, sa taille efficace de population au niveau de la population d'hôtes est égale au produit entre la taille efficace de population dans l'hôte et le nombre d'hôtes. Ainsi, ils affirment que pour des grandes tailles de populations d'hôtes, l'effet de la dérive génétique diminue. Les simulations effectuées avec le modèle épidémiologique présenté dans le chapitre 3 impliquaient 10^3 plantes. Donc même avec une taille efficace de population virale intra-plante de 1 dans les plantes résistantes, d'après Bergstrom et al. (1999) la taille efficace de population au niveau de la

parcelle vaut au moins 10^3 , ce qui doit fortement diminuer l'intensité de la dérive génétique. Pourtant, nous avons pu montrer un effet bénéfique de la dérive génétique. En fait, dans notre modèle, la valeur de la fréquence du variant virulent dans les plantes sensibles, f_{RB} , joue aussi un rôle crucial. Si f_{RB} est très faible, par exemple si elle vaut 10^{-6} , et que la taille efficace de population dans les plantes résistantes N_e^R vaut 1, alors il faudra en moyenne 10^6 tentatives d'infection des plantes résistantes pour qu'une réussisse. Avec 10^3 plantes dans la parcelle, il faudrait donc au moins 10^3 essais d'infection pour chaque plante résistante pour qu'une infection réussisse, et la proportion de plantes sensibles augmente encore ce chiffre. Lorsque f_{RB} est élevée, la probabilité de réussite d'infection d'une plante résistante est forte, et l'effet bénéfique de la dérive génétique est réduit. Je pense que pour les valeurs de f_{RB} les plus faibles testées, nous sommes hors de portée du résultat de Bergstrom et al. (1999), au vu de notre nombre de plantes dans la parcelle. Lorsque f_{RB} est très faible, la diminution de l'effet bénéfique de la dérive génétique rapporté par Bergstrom et al. (1999) devrait être révélée par des nombres de plantes très élevés, et une intensité des épidémies forte.

Par ailleurs, Bergstrom et al. (1999) ont mené leur étude théorique sur un nombre d'hôtes infectés constant. Lorsque des hôtes infectent de nouveaux hôtes, les hôtes sources de l'infection sont retirés de la population, en faisant l'hypothèse de leur mort ou de leur acquisition d'immunité. Dans leur modèle, il n'y a donc pas de dynamique épidémique de type SI (sain - infecté, voir chapitre 2). À l'inverse, nous étudions une dynamique épidémique de type SI. Dans notre modèle, la saison commence avec une seule plante infectée, et nous suivons l'évolution du nombre de plantes infectées qui augmente au cours du temps. Les résultats de Bergstrom et al. (1999) ne s'appliquent donc peut-être pas directement dans notre cas. Notamment, comme le nombre de plantes infectées varie au cours du temps, on pourrait se demander quel est l'impact de cette variation sur la formule de Bergstrom et al. (1999) sur l'intensité de la dérive génétique au niveau d'une population d'hôte. En comparaison à l'approximation par la moyenne harmonique lorsque les tailles efficaces de population varient dans le temps (Caballero, 1994; Kimura, 1970; Kimura and Crow, 1963; Lande and Barrowclough, 1987; Wright, 1939), est-ce que la première phase pendant laquelle peu de plantes sont infectées a plus de poids sur l'intensité de la dérive génétique? Dans ce cas nous aurions un effet moins important du nombre de plantes infectées que celui décrit par Bergstrom et al. (1999). Cela pourrait expliquer pourquoi nous observons un effet bénéfique de la dérive génétique pour des valeurs assez élevées de f_{RB} telle que 10^{-2} (Fig. 3.3).

Par ailleurs, nous supposons que dans les plantes sensibles les deux variants virulent et avirulent arrivent instantanément à l'équilibre sélection-mutation, et que les plantes résistantes

sont instantanément soit infectées par le variant virulent, soit restent saines. À l'inverse, Bergstrom et al. (1999) étudient un changement presque quantitatif (10 changements possibles) d'accumulation de mutations dans la population d'agents pathogènes lors de leur évolution dans les hôtes. Au final, les deux études sont donc assez différentes, Bergstrom et al. (1999) se préoccupant plus de l'évolution intra-hôte, alors que, dans notre cas, nous supposons que l'évolution intra-hôte est fixée instantanément et nous nous préoccupons plus des dynamiques épidémiques. Néanmoins, la taille de la population d'hôte doit avoir un impact sur l'intensité de la dérive génétique à ce niveau, et il serait intéressant de l'étudier.

6.2.2 Dérive génétique et probabilité d'émergence de mutants adaptés aux résistances

Voies de mutation vers la virulence

André and Day (2005) ont étudié la probabilité d'émergence d'agents pathogènes adaptés avant extinction du variant mal adapté au niveau d'une population d'hôtes (humains) (Fig. 6.1). Dans leur modèle le plus simple, ils ont supposé que les infections secondaires arrivaient à un taux constant b sur la durée totale d'une infection. Le taux *per capita* de mortalité des hôtes ou d'élimination de l'agent pathogène par le système immunitaire se passe à un taux d sur toute la durée de l'infection, si bien que la durée attendue d'une infection par le variant mal adapté vaut $L = 1/d$. Le nombre de reproduction de base de l'agent pathogène mal adapté vaut $R_0 = b/d$, avec l'hypothèse $R_0 < 1$. Ce paramètre correspond au nombre moyen d'individus infectés par un individu infecté par l'agent pathogène mal adapté. André and Day (2005) modélisent deux voies distinctes d'adaptation de l'agent pathogène. L'une a lieu lors du goulet d'étranglement de la transmission, pouvant conduire à la fixation d'une mutation adaptative par hasard, avec une probabilité u . L'autre a lieu tout au long de l'infection d'un hôte, au cours de laquelle la population d'agents pathogènes peut changer de génotype à cause de la fixation d'une mutation adaptative, à un taux μ . Lorsqu'une seule mutation est suffisante pour l'adaptation de l'agent pathogène, André and Day (2005) définissent la probabilité d'émergence à partir de variants pathogènes mal adaptés comme valant approximativement :

$$P \approx \frac{1}{1 - R_0} (uR_0 + \mu L) P_a. \quad (6.1)$$

Dans cette équation, le premier facteur $(uR_0 + \mu L)/(1 - R_0)$ correspond à la probabilité qu'une mutation adaptative apparaisse, et le dernier facteur P_a est la probabilité d'épidémie,

sachant que la mutation est apparue. P_a est définie comme $1 - 1/R^*$, avec R^* le nombre de reproduction de base du variant adapté. Cette équation (6.1) suppose des faibles taux de mutation. La figure 6.1 illustre l'effet de R_0 et L sur la probabilité d'émergence P . Ces deux paramètres affectent positivement P . Par contre, individuellement ils n'ont pas un effet important, la probabilité d'émergence ne dépassant par environ $5 * 10^{-2}$ (avec les valeurs de paramètres données dans la légende de la figure 6.1). C'est la combinaison des deux paramètres qui augmente de manière substantielle P , conduisant à un maximum situé vers $P = 0.55$ (pour $R_0 = 0.99$). Ainsi, le variant mal adapté a plus de chances de produire des mutations adaptatives qui conduiront à une épidémie lorsque les infections sont de longue durée (L) et sont bien transmises (R_0).

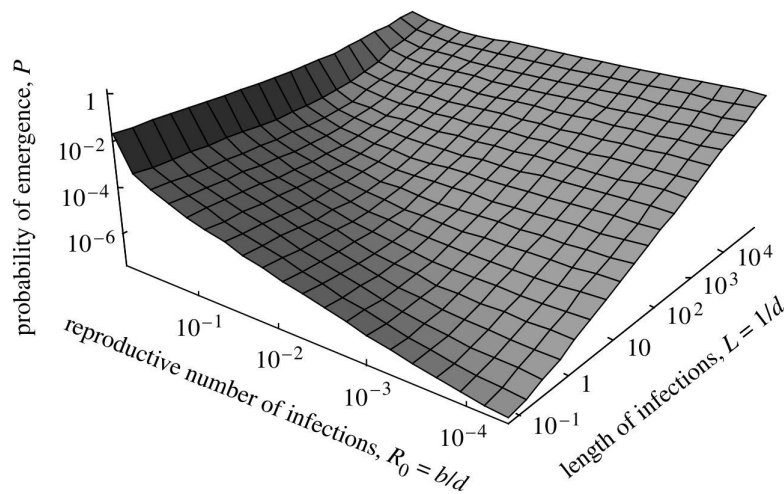


FIGURE 6.1 – **Probabilité P d'émergence d'un mutant adapté en fonction du nombre de reproduction de base R_0 et de la durée attendue L des infections du variant non adapté (équation 6.1).** u (10^{-3}) : probabilité qu'une infection secondaire transmette une mutation adaptative; μ (10^{-6}) : taux d'adaptation pendant la durée d'une infection; P_a (0.5) : probabilité d'émergence de mutants adaptés. Source : (André and Day, 2005).

Dans notre modèle épidémiologique (chapitre 3), le contexte est quelque peu différent car des plantes sensibles sont présentes, pour lesquelles les variants avirulent et virulent sont adaptés. Les plantes sensibles peuvent être vues comme une source d'inoculum continu pour les plantes résistantes, qui ne sont infectées que si le variant virulent survit aux goulets d'étranglement. Dans l'équation 6.1, le paramètre u serait pour nous lié à la probabilité de transmission du variant virulent d'une plante sensible vers une plante résistante, proportionnelle à $1 - e^{-N_e^R f_{RB}}$. Comme dans l'étude de André and Day (2005), le nombre de reproduction de base de l'agent pathogène avirulent (R_0) est inférieur à celui des infections par l'agent pathogène virulent (R^*), l'agent pathogène virulent pouvant infecter les deux variétés sensible et résistante alors que l'agent pathogène avirulent ne peut infecter que les plantes sensibles. Par contre, nous ne

prenons pas en compte la possibilité de mutation du variant avirulent vers le variant virulent dans les plantes résistantes (terme μL dans l'équation 6.1). Dans notre modèle, la durée d'une infection L est fixée par la durée de la saison de culture : nous supposons que chaque plante reste infectée jusqu'à la récolte. Au final, il est intéressant de comparer nos travaux à d'autres, car cela peut apporter des pistes pour des améliorations du modèle. Notamment, on se rend compte ici que l'on pourrait inclure la possibilité de mutation vers le variant virulent dans les plantes résistantes lorsque seul le variant avirulent survit aux goulets d'étranglement.

Durabilité des résistances

Je vais maintenant revenir sur les trois mesures de durabilité d'une résistance au niveau d'un paysage agricole, décrites dans le chapitre 1 d'introduction, proposées par van den Bosch and Gilligan (2003) (Fig. 1.4). Pour rappel, ces auteurs ont différencié deux cas, celui où le variant virulent n'est pas présent initialement dans le paysage (T invasion : temps au bout duquel le variant virulent envahit la population de plantes résistantes, Fig. 1.4), et celui où il est déjà présent (T 'prend le dessus' : temps au bout duquel le variant virulent atteint une certaine fréquence, et T additionnel : nombre d'hôtes sains cumulé dans le temps jusqu'à ce que le variant virulent ait 'pris le dessus', Fig. 1.4). Leurs résultats montrent les différences de durabilité prédites selon la mesure utilisée. Dans l'étude épidémiologique de cette thèse (chapitre 3), la mesure de durabilité est différente puisque le temps est fixé (une saison de culture). On s'intéresse aux gains de rendement additionnels fournis par une stratégie de déploiement d'une variété résistante, calculés à partir des aires sous les courbes d'évolution des proportions de plantes infectées. Ce n'est donc pas non plus une mesure qui étudie directement la fréquence du variant virulent au niveau du paysage, mais elle en est néanmoins indirectement le témoin. De plus, notre étude épidémiologique considère toujours que le variant virulent est initialement présent dans le paysage, et plus précisément dans la plante sensible initialement infectée. Il y est présent à la fréquence f_{RB} correspondant à l'équilibre sélection - mutation, qui dépend du nombre de mutations requises pour le contournement du gène majeur de résistance et de leur coût de fitness dans les plantes sensibles. On peut alors se demander quels seraient les résultats si l'on procédait différemment pour la modélisation des dynamiques génétiques au niveau intra-hôte dans les plantes sensibles, et que le variant virulent soit absent initialement. D'après van den Bosch and Gilligan (2003), les résultats peuvent être différents (Fig. 1.4).

Cela étant dit, les cas où la fréquence du variant virulent dans les plantes sensibles est très faible ($f_{RB} \in [10^{-8}, 10^{-6}]$) pourraient être assimilés à des cas où le variant virulent est

initialement présent en zéro ou une copie tout au plus, si l'on considère par exemple que la taille de population virale maximale dans une plante vaut 10^6 , tel que supposé par Fabre et al. (2012a). Dans ces cas, nos résultats corroborent ceux de van den Bosch and Gilligan (2003), lorsqu'ils passent d'un paysage où le variant virulent est initialement présent à un paysage où il ne l'est pas, à savoir la durabilité mesurée est différente. En ce qui concerne nos résultats, cela se traduit par un seuil de f_{RB} entre les cas où la dérive génétique a globalement peu (ou pas) d'effet sur les gains de rendement ($f_{RB} < 10^{-6}$) et les cas où la dérive génétique peut avoir un effet important ($f_{RB} > 10^{-6}$, Fig. 3.3).

Probabilité de fixation dans l'hôte résistant

En génétique des populations, on retrouve des différences selon les deux conditions initiales - le mutant virulent est initialement présent ou absent - au niveau intra-hôte, pour la probabilité de fixation de mutations bénéfiques (Fig. 6.2) (Barrett and Schluter, 2008; Hermisson and Pennings, 2005). Selon qu'un mutant bénéfique vient d'apparaître en une copie ou qu'il est déjà présent dans la population, les probabilités de fixation de ce mutant peuvent être très différentes en fonction des valeurs de N_e et de s . De manière générale, les probabilités de fixation augmentent avec s et N_e . Lorsque le produit $N_e s$ est faible, les deux probabilités de fixation tendent vers zéro, et lorsqu'il est très élevé, elles tendent vers 1. Pour les cas intermédiaires, la probabilité de fixation d'un mutant apparu en une copie est plus faible que celle d'un allèle déjà présent dans la population. La probabilité de fixation d'un mutant apparu en une copie reste proche de zéro pour une large gamme de valeurs du produit $N_e s$, alors que dans l'autre cas la probabilité de fixation augmente continûment avec $N_e s$ (Fig. 6.2). Cela suggère qu'avec des tailles efficaces de population faibles à modérées et/ou des effets sélectifs plus faibles, les mutations ont davantage de chances de se fixer si elles sont déjà présentes dans la population plutôt que si elles viennent d'apparaître en une copie (Barrett and Schluter, 2008).

Ce dernier cas correspond typiquement au cas d'infections de plantes résistantes. En effet, comme nous l'avons déjà évoqué, les résistances qualitatives ne sont en fait pas totales (Acosta-Leal and Xiong, 2008; Montarry et al., 2011). Notamment, dans les expériences de contournement de la résistance qualitative due à *pvr2³* de Montarry et al. (2011) et Quenouille et al. (2013, 2014) suite à l'inoculation d'un variant avirulent, ce variant avirulent doit vraisemblablement s'être multiplié dans les plantes résistantes pour acquérir la ou les mutation(s) de virulence (Montarry et al., 2011). On peut retrouver la tendance des résultats de Hermisson and Pennings (2005) et Barrett and Schluter (2008) dans notre étude des dégâts relatifs D

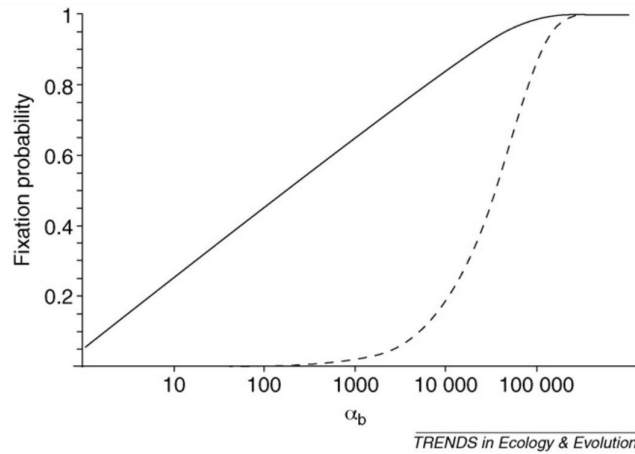


FIGURE 6.2 – Probabilités de fixation d’un mutant apparu en une copie (courbe en pointillés) ou d’un allèle déjà présent dans une population (courbe pleine). $\alpha_b = 2N_e s_b$, où N_e est la taille efficace de la population et s_b est le coefficient de sélection associé à la mutation ou à l’allèle étudié. L’indice b indique que la mutation ou l’allèle sont bénéfiques. L’échelle de α_b est logarithmique. Source : (Barrett and Schluter, 2008; Hermisson and Pennings, 2005).

(ratio de pertes de rendement lorsqu’une variété résistante est déployée par rapport à une parcelle composée uniquement de plantes sensibles, chapitre 3, Fig. 3.4). Dans la figure 3.4, on voit un effet fort de f_{RB} sur les dégâts relatifs, et plus particulièrement un effet seuil de f_{RB} au-delà duquel les dégâts relatifs augmentent jusqu’à atteindre un plateau. Pour des valeurs faibles de f_{RB} ($\leq 10^{-5}$) dans les plantes sensibles, le variant virulent sera transmis en moyenne au plus en une copie aux plantes résistantes. On se trouve ainsi dans le cas de la courbe en pointillés de la figure 6.2, avec une probabilité faible de fixation du variant virulent dans les plantes résistantes, conduisant à des gains de rendement élevés. Pour des valeurs plus élevées de f_{RB} , le variant virulent pourra être transmis en plusieurs copies aux plantes résistantes et on se trouve dans le cas de la courbe pleine de la figure 6.2, avec une probabilité de fixation plus forte, conduisant à des gains de rendement plus faibles. Ensuite, la figure 3.4 montre que la valeur seuil de f_{RB} au-delà de laquelle les dégâts relatifs augmentent dépend de la taille des goulets d’étranglement N_e^R . C’est le produit $f_{RB}N_e^R$ qui détermine le nombre moyen de copies du variant virulent transmises. Quand $f_{RB}N_e^R$ est petit, on se retrouvera plutôt dans le cas de la courbe en pointillés, alors que lorsque $f_{RB}N_e^R$ est grand, on passera à la courbe pleine (Fig. 6.2). À l’extrême, lorsque N_e^R est très grand (α_b très grand dans la figure 6.2), on se rapproche d’un comportement déterministe du système (Rouzine et al., 2001) : la fixation du variant virulent est quasiment certaine, qu’il soit transmis en une copie ou plus. Il est à noter que dans notre modèle, la survie d’une copie du variant virulent conduira de manière certaine à l’infection d’une plante résistante. Le parallèle que je fais avec les travaux de Hermisson and

Pennings (2005) et Barrett and Schluter (2008) concerne donc pour nous la probabilité de survie aux goulets d'étranglement, mais il est intéressant de voir le parallèle qui peut être fait entre notre étude et les leurs.

6.2.3 Pertinence du modèle d'estimation de N_e et s ?

Le modèle LVDM d'estimation de N_e et s présenté dans le chapitre 4 de ce manuscrit a été validé sur une large gamme de valeurs de ces deux paramètres. Un point fort de ce modèle est qu'il ne nécessite pas de marqueurs neutres, contrairement à la large majorité des autres modèles de la littérature estimant N_e et s (Foll et al., 2014b) ou N_e uniquement (Anderson et al., 2000; Berthier et al., 2002; Nei and Tajima, 1981; Waples, 1989; Williamson and Slatkin, 1999).

La critique principale qui peut être émise concernant le modèle LVDM est l'utilisation de formules de débiaisage pour obtenir les estimateurs de N_e et s . En effet, même si ces formules ont été développées avec attention en se basant sur la littérature, nous pouvons nous demander si elles seraient applicables telles quelles dans des cadres expérimentaux différents, c'est-à-dire avec d'autres nombres de variants, d'autres nombres et dates d'échantillonnage, d'autres types d'agents pathogènes ou d'hôtes. Nous avons pu valider le modèle grâce à des simulations numériques du modèle de Wright-Fisher similaires à celles présentées dans le chapitre 4, avec deux variants au lieu de cinq, ou bien une seule date d'échantillonnage au lieu de six, en utilisant les mêmes formules de débiaisage. Pour la validation des estimations dans ces deux cas, nous avons cependant dû définir des critères plus drastiques concernant la diversité minimale restante dans les jeux de données simulés sous FFPopSim (Zanini and Neher, 2012a,b).

L'aspect paraissant le plus critique pour la qualité des estimations est la perte de variants (Luikart et al., 1999; Malaspinas, 2016; Richards and Leberg, 1996). Dans le cas de la validation du modèle avec deux variants, la perte de diversité génétique a un impact majeur à cause du faible nombre de variants, et dans le cas d'une date d'échantillonnage elle est due au fait que l'on observe la composition des populations à une date tardive, et donc qu'il y a davantage de risques pour que certains variants aient disparus de la population par dérive génétique ou par contre-sélection. Cela entraîne un manque d'information quant au moment auquel les variants en question ont été éliminés, et donc une difficulté pour estimer s . Avec leur modèle, Terhorst et al. (2015) rapportent une plus grande capacité à détecter de la sélection lorsque l'intensité de cette sélection est plus élevée, et une plus faible capacité pour un grand nombre de variants. Il y aurait donc un compromis entre un nombre important de variants, conduisant à une plus

forte probabilité de perte de variants, puisqu'ils sont en moyenne présents à des fréquences plus faibles, et un faible nombre de variants, qui peut également rendre délicate voire impossible l'estimation de s si certains de ces variants disparaissent de la population (Malaspinas, 2016). Des simulations effectuées avec huit variants ne m'ont pas permis d'estimer correctement les paramètres N_e et s dans la majorité des cas, lorsque les coefficients de sélection étaient trop faibles (Rousseau et al., Submitted, données non montrées). Pour des estimations correctes, il faudrait un nombre de variants respectant un compromis entre une diversité génétique suffisante sans trop de pertes, et des effets de sélection suffisamment grands pour une estimation plus précise. De plus, des tailles de populations plus grandes permettent d'estimer des coefficients de sélection plus faibles, car l'effet de la dérive génétique en termes de disparition de variants est alors diminué (Malaspinas, 2016). Par ailleurs, Terhorst et al. (2015) rapportent que l'exclusion des variants montrant des fréquences trop proches de 0 ou de 1 dans les analyses permet d'augmenter l'efficacité de leur modèle pour les estimations de N_e et s .

6.2.4 Taille efficace de population et taux d'évolution d'une population : limites actuelles

Le modèle de Wright-Fisher et la formule de probabilité de fixation de Kimura sont, par définition, simplificateurs. Notamment, Lanfear et al. (2014), qui ont étudié le lien entre N_e et le taux d'évolution d'une population à partir de résultats empiriques et théoriques, ont souligné le fait que la relation entre ces deux variables n'était pas toujours aussi simple que dans la théorie. Dans leur revue, le taux d'évolution fait référence au taux auquel de nouvelles mutations qui arrivent à la fixation s'accumulent dans le génome au cours du temps. Pour des mutations affectées d'un coefficient de sélection positif, le taux d'évolution adaptative devrait être positivement corrélé à N_e , et l'effet de N_e devrait être d'autant plus fort que le coefficient de sélection associé aux mutations est plus élevé (Fig. 6.3). Si la majorité des études confirment cette corrélation positive, Lanfear et al. (2014) pointent une étude suggérant un manque de corrélation (droite horizontale) entre le taux d'évolution adaptative et N_e . Une première explication proposée par Lanfear et al. (2014) est que si N_e varie au cours du temps, ce qui peut être le cas pour les virus de plante (Fabre et al., 2012a; Rousseau et al., Submitted), alors différentes mutations vont être soumises à différents N_e , et cela peut conduire à une disparité entre le N_e estimé et le N_e d'intérêt, c'est-à-dire celui qui permettrait de déduire le taux d'évolution de la population. Une deuxième explication concerne le fait que N_e varie le long du génome, ce qui entraînerait des variations dans le taux d'évolution aux sites sous

sélection. Cela rendrait donc plus difficile la détermination du taux d'évolution à l'échelle de la population.

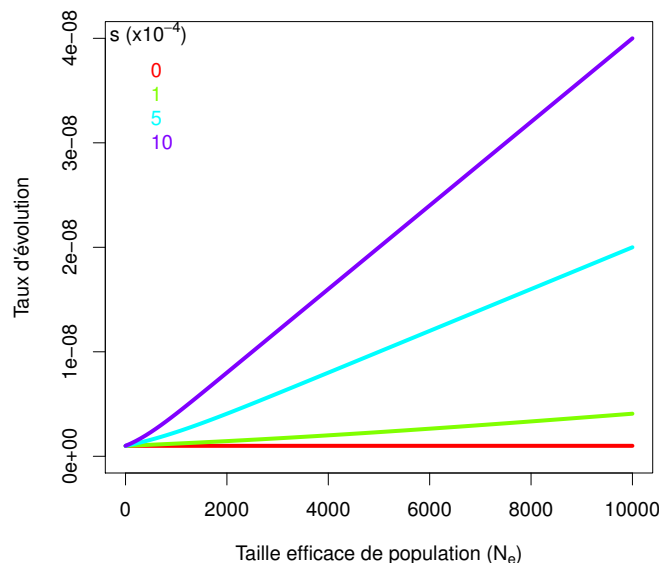


FIGURE 6.3 – **Lien entre le taux d'évolution et la taille efficace d'une population (N_e).** Le taux d'évolution correspond au taux de fixation de mutations par an. Ici, toutes les mutations ont des coefficients de sélection (s) positifs. Source : Lanfear et al. (2014).

6.2.5 Variabilité phénotypique due au génome de l'agent pathogène

Nous nous sommes concentrés sur le génome des plantes, mais un trait phénotypique, qu'il soit qualitatif ou quantitatif, peut être contrôlé par le génome de l'hôte, mais aussi par le génome de l'agent pathogène ainsi que par leur interaction (Fig. 1.7, Lannou (2012)). De manière indirecte, les différences de fitness observées entre variants viraux dans le chapitre 4 révèlent l'importance du génome de l'agent pathogène sur des traits quantitatifs d'intérêt pour prédire le potentiel évolutif de cet agent pathogène face à une résistance. En effet, la fitness d'un variant étant équivalente à son taux de croissance, cela signifie que les différentes mutations propres aux variants étudiés leur confèrent, en plus d'une adaptation à la résistance, un taux de croissance plus ou moins élevé dans les plantes résistantes. Ainsi, le génome des variants viraux ne différant que par ces mutations (G, K et N), cela démontre que le gène codant pour la VPg, où se situent ces mutations, n'a pas uniquement un rôle qualitatif sur la capacité à infecter une plante résistante, mais également un effet quantitatif sur le taux de croissance dans les plantes résistantes. La composition des populations d'agents pathogènes a aussi son importance, et les résultats peuvent diverger d'un pathosystème à un autre. Par exemple, Ayme et al. (2006, 2007) ont montré chez le pathosystème PVY - piment que l'ordre des valeurs de fitness entre variants

était identique, que ces variants soient en compétition ou qu'ils soient seuls dans les plantes. À l'inverse, une expérience similaire de Pierrugues et al. (2007) chez le pathosystème CMV - tabac a montré une tendance opposée en terme de présence de deux variants de CMV dans les feuilles apicales des plantes, lorsque ces variants étaient en infection seuls ou en compétition l'un avec l'autre. De plus, sur nos données, la comparaison de l'ordre des lignées HD pour les coefficients de sélection s entre les variants G et N entre nos expériences de compétition avec deux (données non montrées) ou cinq variants (chapitre 4) a montré que l'ordre de s n'était pas conservé selon le nombre de variants en compétition.

Un effet d'interaction entre l'hôte et son agent pathogène pour l'expression d'un trait quantitatif pourrait modifier le classement des génotypes de plante ou de l'agent pathogène pour ce trait. Par exemple, l'ordre des valeurs de fitness de différents variants pathogènes pourrait être différente selon le génotype de plante. Ce n'est pas ce que l'on observe sur nos données d'estimation des taux intrinsèques d'accroissement naturel (Fig. 4.5). Les autres estimations (N_e , accumulation virale, fréquence de contournement) étant mesurées globalement au niveau de la population virale, je ne peux pas regarder un tel effet pour ces traits. Lê Van et al. (2013) se sont intéressés à de tels effets d'interaction en étudiant les pressions de sélection exercées sur des variants de *Venturia inaequalis* en compétition dans différents génotypes de pommiers, via les proportions de ces différents variants dans les plantes après co-inoculation. Dans certains cas, ils ont pu identifier des QTL dans le génome du pommier contrôlant ces pressions de sélection, associés à chaque fois à un variant de *V. inaequalis*, et ils ont observé le lien entre ces QTL et d'autres contrôlant la sévérité de la maladie (sévérité des sporulations), agissant soit sur un spectre restreint à modéré d'agents pathogènes, soit sur un large spectre d'agents pathogènes. Au final, seuls des QTL de pression de sélection co-localisant avec les QTL de sévérité de la maladie à spectre restreint à modéré ont été détectés. Pour certains variants pathogènes, une réduction de la sévérité de la maladie était associée à une réduction de leur proportion dans les plantes, alors que pour d'autres variants elle était au contraire associée à une augmentation de leur proportion. À l'inverse, aucun QTL de pression de sélection n'a été détecté en co-localisation avec les QTL de sévérité de la maladie agissant à large spectre. Au final, des effets d'interaction entre le génome de l'hôte et de l'agent pathogène apparaissent clairement dans cette étude de Lê Van et al. (2013). Ainsi, il paraît réellement important d'étudier les potentiels effets d'interaction, car de nouveaux variants pathogènes sensibles à de tels effets pourront avoir des conséquences importantes sur l'évolution de la population d'agents pathogènes.

6.2.6 Application à d'autres agents pathogènes ?

Comme nous l'avons vu dans le chapitre 1, l'intérêt de combiner résistances quantitative et qualitative pour protéger une résistance qualitative a été montré sur les virus (Palloix et al., 2009), mais également sur les champignons (Brun et al., 2010) et sur les nématodes (Fournet et al., 2013). À la différence de notre étude sur une résistance quantitative diminuant les tailles efficaces de populations virales dans les plantes, la résistance quantitative étudiée par Brun et al. (2010) diminue la sévérité des symptômes, et celle étudiée par Fournet et al. (2013) diminue la probabilité de développement des nématodes vers le stade adulte mâle. Nous nous sommes concentrés sur la dérive génétique au niveau intra-plante, mais McDonald and Linde (2002a) propose d'autres moyens, au niveau des populations d'hôtes, pour diminuer les tailles efficaces de population, tels que la rotation des cultures, ou de manière naturelle les extrêmes climatiques, notamment la survie à l'hiver. Diminuer la taille réelle de population des agents pathogènes permet de diminuer leur taille efficace de population, ainsi que la probabilité de présence de mutants dans la population, ce qui doit diminuer le risque d'émergence de variants virulents (McDonald and Linde, 2002a; Zhan et al., 2015). Dans leur revue, McDonald and Linde (2002a) listent plusieurs pathosystèmes impliquant des champignons pour lesquels les tailles efficaces de population au niveau des cultures sont petites à intermédiaires. Au niveau des populations d'hôtes, les solutions évoquées s'appliquent à tous les agents pathogènes. Au niveau d'un hôte, il serait intéressant d'étudier plus en détails si des mécanismes génétiques peuvent permettre de réduire les tailles de populations d'agents pathogènes autres que les virus. Si c'est le cas, alors les études présentées dans ce manuscrit pourraient s'appliquer à ces pathosystèmes. À l'intersection entre les niveaux intra-hôte et inter-hôte, rappelons que, d'après Bergstrom et al. (1999), des tailles de populations d'hôtes plus petites permettent d'augmenter l'intensité de la dérive génétique au niveau de la population d'hôte.

Dans notre cas, au niveau intra-hôte nous avons pu estimer séparément taille efficace de population et taille réelle de population (via l'accumulation virale). Même si ces deux variables sont, au moins aux dates précoces de l'infection, assez bien corrélées (chapitre 5), la corrélation n'est pas parfaite et il est peut-être possible de jouer séparément sur les deux, dans une certaine mesure. Il pourrait être intéressant d'approfondir cet aspect, en étudiant si les deux peuvent être exploités séparément et quel en serait l'intérêt. Notamment, la taille efficace de population détermine l'intensité de la dérive génétique, et la taille réelle de population peut être vue comme un proxy de la probabilité de présence de mutants dans la population. Il reste à étudier si la possibilité d'agir séparément sur les deux forces évolutives de dérive génétique et de mutation

pourrait permettre un meilleur contrôle des populations virales.

6.3 Perspectives

6.3.1 Identifier des QTL de taille efficace de populations virales chez les plantes hôtes

Potentiellement, il serait envisageable d'identifier des QTL de plantes contrôlant une part significative de la variabilité des traits de tailles efficaces de populations virales à différentes étapes du cycle infectieux, et de les prendre en considération pour la sélection de variétés porteuses de résistances plus durables. Une partie de mes travaux de thèse, que je n'ai pas présentée dans ce manuscrit, a consisté à identifier des QTL de tailles efficaces de population à l'inoculation. Nous avons ainsi détecté trois QTL chez le piment, pour chacun des deux virus PVY et CMV, expliquant respectivement 57.82% et 50.88% de la variation du trait phénotypique étudié (effets additifs et épistatiques entre paires de loci) (Tamisier et al., 2016). Des goulets d'étranglement étant également présents à d'autres moments de l'infection d'une plante, il serait maintenant intéressant de tenter d'identifier des QTL pour ces autres goulets. Un des buts de l'expérience 5 du chapitre 5 (Table 5.1), estimant N_e et s en infection systémique sur 90 lignées HD, était aussi d'identifier des QTL correspondant à ces deux traits phénotypiques. Cependant, pour l'instant un seul QTL a pu être détecté, expliquant une part de la variabilité de s . Cela peut être imputable à un échantillonnage à une seule date tardive qui a pu diminuer la fiabilité des estimations à cause d'un nombre réduit de données et d'une perte parfois importante de diversité génétique dans les populations virales analysées. De plus, les résultats du chapitre 4 ont montré que les tailles efficaces de population aux dates précoces sont davantage corrélées à la fréquence de contournement que celles estimées aux dates tardives. Ainsi, un échantillonnage effectué plus tôt, par exemple dans la feuille inoculée vers 5 jours après inoculation (jai), et/ou au début de l'infection systémique vers 10 jai, aurait peut être permis une identification de QTL. Néanmoins, l'importante variabilité observée pour ces traits est déjà suffisante pour identifier des lignées HD permettant de diminuer de manière conséquente les tailles efficaces de populations virales. Ces caractéristiques pourraient être intégrées dans les programmes d'élaboration de nouvelles variétés.

6.3.2 Modèle d'estimation de N_e et s

Pour revenir au modèle LVDM d'estimation de N_e et s , la validité des formules de débiaisage pour les deux cas supplémentaires testés, à savoir soit deux variants soit une date, sont des bons indicateurs de leur potentielle large applicabilité. Néanmoins, d'autres tests demanderaient à être effectués, notamment sur les critères d'acceptabilité des jeux de données en fonction du nombre de variants et des dates d'échantillonnage. Cela permettrait d'évaluer les gammes de valeurs de N_e et s pour lesquelles le compromis discuté entre nombre de variants et intensités de la sélection et de la dérive génétique (Terhorst et al., 2015) rend les estimations acceptables ou non. Enfin, une piste pour éviter une trop forte diminution de diversité génétique, qui compromet la qualité des estimations (Luikart et al., 1999; Malaspinas, 2016; Richards and Leberg, 1996), serait d'adapter les dates d'échantillonnage selon la vitesse d'évolution des populations. Cette vitesse d'évolution est liée aux intensités de sélection et de dérive génétique, mais aussi au temps de génération. Pour une population évoluant rapidement, il faudrait ainsi privilégier des échantillonnages à des dates proches les unes des autres pour éviter trop de pertes de variants.

6.3.3 Améliorations du modèle épidémiologique

Dans la discussion du chapitre 3, plusieurs critiques ont été formulées quant au modèle épidémiologique proposé. Je vais ici approfondir les points qui me paraissent les plus importants.

La dérive génétique est-elle toujours bénéfique pour la durabilité des résistances ?

Dans notre modèle épidémiologique, la dérive génétique a toujours un effet bénéfique ou neutre sur les gains de rendement additionnels, mais jamais un effet négatif notable. La meilleure stratégie est toujours de réduire au maximum les tailles de goulet d'étranglement subis par les populations virales. Or, dans la littérature, Miyashita and Kishino (2010), Zwart and Elena (2015) et Levy et al. (2015) évoquent des situations où des goulets d'étranglement étroits pourraient conduire à une adaptation plus rapide. Miyashita and Kishino (2010) et Zwart and Elena (2015) se focalisent sur des résistances qui nécessitent plusieurs mutations pour être contournées, et dont les mutants intermédiaires ont une fitness plus faible que le variant avirulent de départ ainsi que le variant virulent. Ils expliquent que des goulets d'étranglement très étroits pourraient permettre, par un effet d'échantillonnage aléatoire, d'isoler les variants intermédiaires, et ainsi leur donner une chance plus importante d'acquérir les autres mutations pour la virulence. Typiquement, lors de la colonisation des cellules d'une feuille, les variants

intermédiaires peuvent se retrouver isolés dans une cellule. De leur côté, Levy et al. (2015) expliquent que les réductions saisonnières drastiques des populations de cochons d'Inde au Pérou, hôtes de *Trypanosoma cruzi*, l'agent de la maladie de Chagas, sont responsables du maintien de cette maladie. Les auteurs pensent que ces réductions de tailles de populations d'hôtes entraînent une augmentation de la prévalence de *T. cruzi* dans les insectes vecteurs, qui s'alimentent alors sur un nombre réduit d'hôtes et ont donc une plus forte probabilité de s'alimenter sur des hôtes infectés. En retour, cela accroît le risque de maladie dans les maisons. Ces cas de figures, même s'ils sont éloignés de notre cadre d'études, montrent tout de même que les goulets d'étranglement, parce qu'ils correspondent à un échantillonnage aléatoire des individus dans une population, peuvent aussi conduire à la sélection de variants adaptés et donc dans notre cas conduire à une adaptation rapide des virus à la résistance des plantes. Pour revenir aux remarques de Miyashita and Kishino (2010) et Zwart and Elena (2015), Park et al. (2013) ont étudié l'effet du positionnement de la fitness d'un variant intermédiaire par rapport au variant initial et au variant final sur la probabilité d'émergence du variant final au niveau d'une population d'hôtes. Ils ont montré que la probabilité d'émergence augmentait avec la fitness du variant intermédiaire, cette probabilité étant la plus faible lorsque la fitness du variant intermédiaire était inférieure aux deux autres variants, et étant la plus forte lorsque la fitness du variant intermédiaire était entre les fitness des deux autres variants.

Dans notre modèle épidémiologique, l'hypothèse selon laquelle les populations virales arrivent instantanément à leur équilibre dans les plantes peut, en effet, être simpliste. Le relâchement de cette hypothèse permettrait peut être de voir les effets néfastes possibles de la dérive génétique. Notamment, Gerrish and Lenski (1998) et Sniegowski and Gerrish (2010) avancent que lorsque la taille efficace de population est grande, le temps à la fixation est long mais la probabilité de fixation est forte, alors que lorsque la taille efficace de population est petite, le temps à la fixation est court mais la probabilité de fixation est faible. Ainsi, nous pourrions obtenir un modèle plus proche de la réalité en introduisant un compromis entre la probabilité de fixation du variant virulent dans les plantes résistantes et le temps pour la fixation de ce variant. Cela pourrait être mis en place assez simplement, en introduisant un nouveau compartiment dans le modèle, celui des plantes exposées E. Les plantes resteraient dans ce compartiment plus ou moins longtemps selon que les tailles efficaces de populations virales qu'elles induisent sont grandes ou petites. Un seul taux de transition supplémentaire serait suffisant, définissant la sortie d'une plante du compartiment E, soit pour retourner vers le compartiment des plantes saines S, soit pour aller dans le compartiment des plantes infectées I, selon un tirage dans une loi de Bernoulli ayant pour paramètre la probabilité de fixation

du variant virulent. Au final, cette modification permettrait de pénaliser le virus dans les cas où les tailles efficaces de populations virales sont grandes en retardant l'infection des plantes résistantes, ce qui devrait retarder les épidémies dans ces cas-là.

De plus, dans le chapitre 3 nous nous sommes principalement focalisés sur les valeurs moyennes de gains de rendement additionnels. Or, la variabilité dans les résultats des simulations, due à la dérive génétique et à la stochasticité des transmissions, peut avoir des conséquences très diverses sur la durabilité d'une résistance. Lors d'une saison de culture, ce n'est pas le comportement moyen simulé mais bien celui d'une simulation en particulier qui se passe. Ainsi, il paraîtrait important d'approfondir les analyses en étudiant également la variabilité des dynamiques épidémiques selon les valeurs de paramètres, par exemple grâce à la distribution de probabilité des gains de rendement au dernier jour de la saison de culture. Peut être identifierons-nous plus de cas où la dérive génétique peut avoir un effet néfaste pour la durabilité des résistances, à cause d'une infection précoce de la première plante résistante par un effet aléatoire lié à la fixation rapide d'un variant virulent dans une plante.

Relâchement du lien entre proportion de plantes résistantes et intensité de la dérive génétique

À cause du pyramidage d'une résistance qualitative avec une résistance quantitative contrôlant les tailles de goulets d'étranglement, nous avons forcé un lien entre la proportion de plantes résistantes dans la parcelle et l'intensité de la dérive génétique. Cela a conduit à préférer des proportions importantes de plantes résistantes pour diminuer les pertes de rendement, alors que d'autres études ont montré que des proportions intermédiaires de plantes résistantes étaient la plupart du temps optimales (Fabre et al., 2012b; Ohtsuki and Sasaki, 2006). Aussi, nous avons simulé les épidémies sur une saison de culture, donc nous nous sommes plus focalisés sur les gains de rendements à court terme que sur la durabilité de la résistance à long terme. Cependant, même si Fabre et al. (2012b) ont simulé les épidémies sur quinze ans, Ohtsuki and Sasaki (2006) les ont simulé sur un an, comme nous. Je me demande quels seraient les résultats si nous introduisions quatre variétés de plantes : des plantes sans gène majeur de résistance et sans résistance quantitative (S_s), des plantes sans gène majeur mais avec résistance quantitative (S_r), et des plantes avec gène majeur, soit sans résistance quantitative (R_s), soit avec (R_r). Cela permettrait de dissocier proportion de plantes résistantes et intensité de la dérive génétique. Les plantes S_s pourraient avoir leur importance, constituant un réservoir de variants avirulents, diluant ainsi la proportion du variant virulent au niveau du paysage.

Érosion de la résistance quantitative

De plus, une modification intéressante du modèle serait d'intégrer la possible érosion de la résistance quantitative, telle qu'étudiée par Montarry et al. (2012). Cela implique le suivi de l'évolution d'un trait quantitatif de l'agent pathogène, en l'occurrence ici celui de l'évolution des tailles de goulets d'étranglement. Cette augmentation serait cette fois-ci contrôlée par le génome de l'agent pathogène (Lannou, 2012). Cette modification peut typiquement être mise en place par l'intégration d'une nouvelle variable d'état dans le système, à savoir la taille des goulets d'étranglement. Notamment, Day and Proulx (2004) fournissent une théorie générale pour intégrer l'évolution moyenne d'un trait quantitatif. Ils prennent l'exemple de l'agressivité, mais leur théorie peut s'appliquer à d'autres traits. Nous pourrions nous inspirer de cette théorie pour autoriser les virus à s'adapter à la résistance quantitative.

De manière intéressante, van den Berg et al. (2014) ont montré que la présence d'un QTL pouvait conduire à des changements dans des traits quantitatifs d'un agent pathogène, pas forcément directement liés à celui ciblé par le QTL. Plus précisément, pour compenser l'effet d'un QTL sur un trait quantitatif, l'agent pathogène peut évoluer vers une modification d'un autre trait, qui aura le même résultat final. Les auteurs prennent l'exemple du taux de production de spores par lésion, affecté par la présence de QTL. L'agent pathogène peut évoluer pour produire des lésions de plus grande surface, trait qui n'est pas directement visé par le QTL, mais qui conduira au final à une augmentation de la production de spores par lésion. Dans notre cas, nous pourrions par exemple faire le lien entre N_e et l'accumulation virale. On s'intéresse à des QTL qui diminuent N_e , mais si le virus évolue vers de plus fortes accumulations, cela devrait aussi augmenter sa taille efficace de population, puisque N_e représente une proportion de la population. Aussi, nous avons vu l'importance du produit $N_e s$ sur l'évolution d'une population (Charlesworth, 2009). Ainsi, en présence de QTL contrôlant N_e , si l'agent pathogène évolue vers des valeurs de s plus grandes, alors il pourra contre-balancer l'effet de la dérive génétique en augmentant l'efficacité de la sélection ainsi que la probabilité de fixation d'une mutation bénéfique (Fig. 1.13) (Kimura, 1955, 1962, 1970). Les interactions sont importantes lorsque l'on étudie l'effet de forces évolutives, donc l'effet pointé par van den Berg et al. (2014) peut être observé, et par conséquent il serait intéressant de l'étudier.

Accroître les échelles de temps et d'espace

Pour une étude à proprement parler de la durabilité d'une résistance, il faudrait élargir les simulations sur plusieurs saisons de cultures. Aussi, un réservoir constitué de plantes sauvages

peut être modélisé, hébergeant les agents pathogènes d'une saison de culture à la suivante, responsable de l'initiation des épidémies au début d'une saison, et générant un flux constant de virus vers les cultures. De plus, pour être efficace, une stratégie de contrôle doit être pensée à la même échelle que celle à laquelle se déroulent les épidémies (Dybiec et al., 2004; Gilligan, 2008). Il serait donc intéressant d'étudier les épidémies dans un paysage, constitué de plusieurs parcelles connectées entre elles par des flux de vecteurs. J'ai déjà contribué à un tel type de modélisation, sur plusieurs saisons de cultures et dans un paysage agricole constitué des parcelles cultivées et d'un réservoir (Fabre et al., 2012b, 2015). La structure de ce modèle pourrait être reprise pour étudier l'effet de la dérive génétique sur la durabilité de la résistance des plantes.

6.4 Remarques finales

Dans leur revue, REX Consortium (2013) indiquent le manque d'études sur l'adaptation des populations d'agents pathogènes ou de ravageurs aux pesticides ou aux médicaments, dans le cas de populations de taille finie, donc avec de la stochasticité dans l'évolution des populations d'agents pathogènes. Ils citent seulement deux études qui prennent cet aspect en compte. Plus récemment, Abel et al. (2015) ont pointé du doigt une question d'intérêt qui demande à être étudiée : Pouvons-nous tirer profit des goulets d'étranglement pour concevoir des meilleurs traitements d'administration d'antimicrobiens ou de vaccination, qui limitent l'émergence de variants adaptés et renforcent l'efficacité thérapeutique ? C'est précisément le sujet de l'étude épidémiologique du chapitre 3. Elle a été conduite sur les virus dans les cultures, mais le principe doit *a priori* rester le même pour la vaccination ou la prise d'antimicrobiens. Notre modèle donne donc une première réponse à une question d'actualité.

Concernant les agents pathogènes des cultures, Zhan et al. (2014, 2015) discutent de l'intérêt de la dérive génétique pour une meilleure gestion de la résistance des plantes. Ils proposent plusieurs solutions pour augmenter l'intensité de la dérive génétique : de longues périodes de jachère entre les saisons de culture, des pratiques prophylactiques dans les parcelles (brûlis ou enterrement des résidus de cultures, éliminations des réservoirs ou des plantes infectées pendant la saison de culture), les mélanges de variétés et la rotation des cultures. De manière originale, la solution proposée dans ce manuscrit va dans le sens d'une augmentation de la dérive génétique, mais elle est différente de celles proposées par Zhan et al. (2014, 2015), puisqu'elle considère l'utilisation de résistances quantitatives dans le génome des plantes diminuant les tailles efficaces de populations d'agents pathogènes.

La dérive génétique n'est pas la seule force évolutive, et d'autres peuvent être intéressantes pour une meilleure gestion de la résistance des plantes. Notamment, l'accumulation virale, proxy de la probabilité de présence d'un mutant dans la population, peut aussi constituer un bon levier d'action. On ne l'a pas identifié, mais d'après la théorie de génétique des populations, le coefficient de sélection doit aussi jouer sur la probabilité de fixation des mutations de virulence (Kimura, 1955, 1962, 1970). Cependant, dans le cas des résistances qualitatives que nous étudions, le coefficient de sélection doit être très grand entre les variants avirulent et virulent dans les plantes résistantes. Étant donné que nous étudions des variétés sensible et résistante en mélange au niveau d'une population d'hôtes, il serait certainement judicieux de plutôt s'orienter vers des variétés sensibles conduisant à des coefficients de sélection plus grands entre les variants virulent et avirulent. En effet, dans ces plantes le variant virulent est généralement contre-sélectionné, et nous avons montré l'importance de la fréquence du variant virulent dans les plantes sensibles sur les dynamiques épidémiques.

Dans le chapitre 1 d'introduction, j'avais évoqué d'autres solutions pour une meilleure gestion de la résistance des plantes. L'une d'entre elles consiste à pyramider des résistances qualitatives (McDonald and Linde, 2002a; Mundt, 2002; Pedersen and Leath, 1988; Pink, 2002). Pink (2002) rapporte qu'il était difficile de pyramider des gènes de résistances non contournés avant la naissance de la sélection variétale assistée par marqueur génétique, car aucune méthode ne permettait de discriminer entre une variété portant un gène de résistance ou plusieurs. Ainsi, les sélectionneurs reprenaient une variété portant un gène de résistance qui avait été contourné pour lui introgresser un nouveau gène de résistance. Dans de telles conditions, la durabilité des résistances n'est pas vraiment améliorée. Le développement de la sélection assistée par marqueur a permis de pyramider des gènes de résistances encore efficaces. D'après McDonald and Linde (2002a), ce pyramidage peut être favorable à une plus grande durabilité des résistances car il est alors plus difficile pour l'agent pathogène d'accumuler toutes les mutations nécessaires au contournement des résistances. Une étude expérimentale au champ sur *Phytophthora infestans* s'est intéressée à tester toutes les variétés possibles de pomme de terre en combinant deux gènes majeurs : une variété sensible, une variété avec un gène de résistance, une autre avec le deuxième gène de résistance, et une quatrième variété avec les deux gènes de résistance (Tan et al., 2010). Leur étude a révélé un effet additif des deux gènes de résistance combinés sur la durabilité des résistances, montrant l'intérêt d'une telle combinaison. Cependant, on aurait pu s'attendre à un effet plus qu'additif, ce qui n'est pas le cas dans cette étude.

Une autre solution consiste à pyramider des résistances quantitatives uniquement. Récem-

ment, Bourget et al. (2015) ont conçu un modèle stochastique décrivant les dynamiques évolutives d'un agent pathogène face à des résistances quantitatives au niveau d'un hôte. Ils ont étudié des résistances quantitatives affectant deux traits phénotypiques : le taux de reproduction de l'agent pathogène et sa capacité à coloniser les plantes. Les valeurs des traits sont diminuées par un certain facteur dans les plantes résistantes par rapport aux plantes sensibles, mais l'agent pathogène peut acquérir des mutations par stochasticité, et avec l'effet de la sélection ces mutations peuvent se propager dans la population d'agents pathogènes et conduire ainsi à une augmentation des valeurs des traits affectés. Bourget et al. (2015) ont testés plusieurs scénarios d'adaptation de l'agent pathogène : (i) la restauration identique, conduisant à l'augmentation d'un facteur identique des deux traits suite à l'acquisition d'une mutation, (ii) la restauration spécifique, conduisant à l'augmentation de la valeur d'un seul des deux traits, l'autre trait restant à une valeur constante sur toute la simulation, (iii) la restauration indépendante, conduisant à une augmentation de la valeur d'un trait ou de l'autre au cours de la simulation, et (iv) la restauration antagoniste, lorsque l'augmentation de la valeur d'un trait s'accompagne de la diminution de la valeur de l'autre trait (compromis entre les deux traits). Au final, leur modèle suggère une plus grande durabilité des résistances si l'évolution des traits affectés est antagoniste. Sans une telle caractéristique, Bourget et al. (2015) concluent à une plus grande durabilité de la résistance affectant le taux de reproduction seule. Cette étude théorique est complémentaire de celle de van den Berg et al. (2014) présentée précédemment, mettant en garde sur l'évolution possible d'un trait non lié à la résistance quantitative déployée, mais qui aura le même effet final que si l'agent pathogène avait évolué au trait affecté. Ces résultats suggèrent que des études approfondies sont nécessaires lorsque l'on souhaite utiliser plusieurs résistances quantitatives conjointement.

Gandon and Michalakis (2000) ont comparé l'effet du déploiement d'une résistance qualitative ou quantitative, affectant le taux de croissance de l'agent pathogène dans l'hôte, sur l'évolution de l'agressivité de l'agent pathogène. Leur modèle a montré que le renforcement d'une résistance qualitative, via une augmentation de la proportion de plantes résistantes, conduisait à une diminution de l'agressivité, alors que le renforcement d'une résistance quantitative conduisait au contraire généralement à une augmentation de l'agressivité. Leurs résultats corroborent ceux de May and Nowak (1994) sur les résistances qualitatives et ceux de van Baalen (1998) sur les résistances quantitatives. Ainsi, l'étude de la durabilité n'est pas le seul critère important, l'étude de l'évolution de l'agressivité de l'agent pathogène face aux résistances devrait également être prise en compte, dans des modèles mettant en commun ces deux aspects. Au final, l'étude de Gandon and Michalakis (2000) montre qu'il est préférable de déployer des résistances

qualitatives. Leur étude ne dit cependant rien des déploiements conjoints de résistances qualitative et quantitative. On peut donc se demander quel est l'impact d'un tel déploiement sur l'évolution de l'agressivité de l'agent pathogène.

Pour revenir aux combinaisons de résistances qualitative et quantitative, la plupart des études montrent un effet bénéfique pour la durabilité de la résistance qualitative (Brun et al., 2010; Fournet et al., 2013; Palloix et al., 2009), incluant les études présentées dans ce manuscrit (chapitres 3 et 5). Cependant, Barbary et al. (2014) citent deux études expérimentales rapportant que le pyramidage entre un gène majeur de résistance et des QTL n'a pas d'effet notable (Riedel et al., 2011; Tan et al., 2009). Tan et al. (2009) se sont attachés à étudier l'effet de la combinaison d'un QTL majeur (fournissant un niveau élevé de résistance) avec un QTL mineur chez la pomme de terre sur la réduction du nombre d'œufs développés de *Meloidogyne hapla*. Alors que le QTL majeur réduit drastiquement le nombre d'œufs développés (88%), le QTL mineur ne fournit pas une réduction significative (55%), et la combinaison des deux QTL ne conduit pas à une réduction plus importante que lorsque le QTL majeur seul est déployé. Riedel et al. (2011) ont étudié l'effet de la combinaison de deux gènes majeurs de résistance et d'un QTL chez l'orge sur le *Barley yellow dwarf virus* (genre *Luteovirus*) (BYDV). Alors que les deux gènes majeurs de résistance seuls conféraient une tolérance au BYDV, leur combinaison a procuré une diminution significative de l'accumulation virale ainsi que des gains de rendement plus importants. Le QTL n'a montré aucun effet significatif pour la tolérance, qu'il soit introgressé avec chacun des gènes majeurs seuls ou avec les deux gènes majeurs pyramidés. Par contre, ce QTL a bien un effet significatif dans des plantes ne portant aucun des deux gènes majeurs. Au final, les études de Tan et al. (2009) et Riedel et al. (2011) s'inscrivent dans un contexte assez différent du nôtre, puisque les résistances étudiées agissent à la base sur le même trait de vie de l'agent pathogène, à savoir le nombre d'œufs développés (Tan et al., 2009) ou la tolérance au virus (Riedel et al., 2011). À l'inverse, nos travaux se concentrent sur des résistances qui agissent sur des traits différents : la capacité à infecter les plantes porteuses du gène majeur de résistance, la taille efficace de population virale, ainsi que le coefficient de sélection et l'accumulation virale. Il serait intéressant d'étudier l'effet de la combinaison de résistances quantitatives agissant sur ces différents traits sur les valeurs associées à ces traits. Est-ce que de telles combinaisons entraînent un effet moins qu'additif, additif, ou plus qu'additif ?

Au final, il me semble intéressant d'exploiter le fait que des résistances quantitatives contrôlant différents traits de vie aient été identifiés dans de nombreux pathosystèmes. Nous avons montré que l'ajout d'une résistance quantitative augmentant l'intensité de la dérive génétique à une résistance qualitative permettait d'augmenter les gains de rendement et de

protéger la résistance qualitative. Il paraît donc intéressant de combiner les effets de différents facteurs pour une meilleure gestion finale. Pour la suite de mes travaux, je pense qu'une perspective importante serait d'étudier l'effet des combinaisons entre résistances quantitatives agissant sur différents traits, toujours dans une perspective de plus grande durabilité d'une résistance qualitative. On peut se demander quelles résistances quantitatives sont les plus bénéfiques, lesquelles il ne faut pas associer, et lesquelles au contraire il pourrait être d'autant plus intéressant d'associer ?

Bibliographie

- S. Abel, P. Abel zur Wiesch, B. M. Davis, and M. K. Waldor. Analysis of bottlenecks in experimental models of infection. *PLoS Pathog*, 11(6) :e1004823, 2015. doi : 10.1371/journal.ppat.1004823.
- R. Acosta-Leal and Z. Xiong. Complementary functions of two recessive r-genes determine resistance durability of tobacco 'Virgin A Mutant' (VAM) to *potato virus y*. *Virology*, 379 : 275–283, 2008. doi : 10.1016/j.virol.2008.06.026.
- G. Agrios. *Plant Pathology*. Elsevier Academic, 2005.
- Z. Agur, L. Cojocaru, G. Mazor, R. M. Anderson, and Y. L. Danon. Pulse mass measles vaccination across age cohorts. *Proc. Natl. Acad. Sci. USA*, 90 :11698–11702, 1993.
- E. C. Anderson, E. G. Williamson, and E. A. Thompson. Monte Carlo Evaluation of the Likelihood for N_e From Temporal Spaced Samples. *Genetics*, 156 :2109–2118, 2000.
- P. K. Anderson, A. A. Cunningham, N. G. Patel, F. J. Morales, P. R. Epstein, and P. Daszak. Emerging infectious diseases of plants : pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution*, 19(10) :535–544, 2004.
- R. M. Anderson and R. M. May. Population biology of infectious diseases : Part i. *Nature*, 280 (2) :361–367, 1979.
- J.-B. André and T. Day. The effect of disease life history on the evolutionary emergence of novel pathogens. *Proc. R. Soc. B*, 272 :1949–1956, 2005. doi : 10.1098/rspb.2005.3170.
- R. Antia, R. R. Regoes, J. C. Koella, and C. T. Bergstrom. The role of evolution in the emergence of infectious diseases. *Nature*, 426 :658–661, 2003. doi : 10.1038/nature02104.
- S. Astier, J. Albouy, Y. Maury, and H. Lecoq. *Principes de virologie végétale*. Mieux Comprendre, 2001.

- V. Ayme, S. Souche, C. Caranta, M. Jacquemond, J. Chadœuf, A. Palloix, and B. Moury. Different Mutations in the Genome-Linked Protein VPg of *Potato virus Y* Confer Virulence on the *pvr2*³ Resistance in Pepper. *MPMI*, 19(5) :557–563, 2006.
- V. Ayme, J. Petit-Pierre, S. Souche, A. Palloix, and B. Moury. Molecular dissection of the potato virus Y VPg Virulence factor reveals complex adaptations to the *pvr2* resistance allelic series in pepper. *Journal of General Virology*, 88 :1594–1601, 2007.
- D. J. Balding. Likelihood-based inference for genetic correlation coefficients. *Theor. Pop. Biol.*, 63 :221–230, 2003.
- D. J. Balding and R. A. Nichols. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and patternity. *Genetica*, 96 :3–12, 1995.
- E. Ballini, J. B. Morel, G. Droc, A. Price, B. Courtois, J.-L. Nottoghem, and D. Tharreau. A Genome-Wide Meta-Analysis of Rice Blast Resistance Genes and Quantitative Trait Loci Provides New Insights into Partial and Complete Resistance. *MPMI*, 21(7) :859–868, 2008.
- A. Barbary, A. Palloix, A. Fazari, N. Marteu, P. Castagnone-Sereno, and C. Djian-Caporalino. The plant genetic background affects the efficiency of the pepper major nematode resistance genes *Me1* and *Me3*. *Theor Appl Genet*, 127 :499–507, 2014.
- J. A. Barrett. Pathogen evolution in multiline and variety mixtures. *Journal of Plant Diseases and Protection*, 87(7) :383–396, 1980.
- R. D.H. Barrett and D. Schluter. Adaptation from standing genetic variation. *TRENDS in Ecology and Evolution*, 23(1) :38–44, 2008.
- M. A. Beaumont and D. J. Balding. Identifying adapting genetic divergence among populations from genome scans. *Molecular Ecology*, 13 :969–980, 2004.
- M. Begon, C. R. Townsend, and J. L. Harper. *Ecology : From Individuals to Ecosystems, 4th Edition*. Blackwell Publishing Ltd, 2006.
- C. T. Bergstrom, P. McElhany, and L. A. Real. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proc. Natl. Acad. Sci. USA*, 96 :5095–5100, 1999.
- D. Bernoulli and S. Blower. An attempt at a new analysis of the mortality caused by smallpox and of advantages of inoculation to prevent it. *Rev. Med. Virol.*, 14(5) :275–288, 2004.

- D. Bertacchi, F. Zucca, F. Foresti, D. Mangioni, and A. Gori. Combination versus sequential monotherapy in chronic HBV infection : a mathematical approach. *Mathematical Medicine and Biology*, 32 :383–403, 2015.
- P. Berthier, M. A. Beaumont, J.-M. Cornuet, and G. Luikart. Likelihood-Based Estimation of the Effective Population Size Using Temporal Changes in Allele Frequencies : A Genealogical Approach. *Genetics*, 160 :741–751, 2002.
- M. Betancourt, A. Fereres, A. Fraile, and F. García-Arenal. Estimation of the Effective Number of Founders That Initiate an Infection after Aphid Transmission of a Multipartite Plant Virus. *J. Virol.*, 82(24) :12416–12421, 2008.
- J. P. Bollback, T. L. York, and R. Nielsen. Estimation of $2N_e s$ from temporal allele frequency data. *Genetics*, 179 :497–502, 2008.
- R. Bommarco, D. Kleijn, and S. G. Potts. Ecological intensification : harnessing ecosystem services for food security. *Trends in Ecology and Evolution*, 28(4) :230–238, 2013.
- I. M. Bomze. Lotka-Volterra Equation and Replicator Dynamics : A Two-Dimensional Classification. *Biol. Cybern.*, 48 :201–211, 1983.
- R. Bourget, L. Chaumont, C.-E. Durel, and N. Sapoukhina. Sustainable deployment of QTLs conferring qualitative resistance to crops : first lessons from a stochastic model. *New Phytologist*, 206 :1163–1171, 2015.
- F. Brauer and C. Castillo-Chávez. *Mathematical Models in Population Biology and Epidemiology*. Springer, 2001.
- F. Brauer, P. van den Driessche, and J. Wu. *Mathematical Epidemiology*. Springer, 2008.
- M. A. Brockhurst, N. Colegrave, and D. E. Rozen. Next-generation sequencing as a tool to study microbial evolution. *Molecular Ecology*, 20 :972–980, 2011.
- J. K. M. Brown. Durable Resistance of Crops to Disease : A Darwinian Perspective. *Annu. Rev. Phytopathol.*, 53 :24.1–24.27, 2015.
- J. K. M. Brown and M. S. Hovmøller. Aerial Dispersal of Pathogens on the Global and Continental Scales and Its Impact on Plant Disease. *Science*, 297 :537–541, 2002.
- J. K. M. Brown and A. Tellier. Plant-Parasite Coevolution : Bridging the Gap between Genetics and Ecology. *Annu. Rev. Phytopathol.*, 49 :345–67, 2011.

- J. K. M. Brown, M. S. Hovmøller, R. A. Wyand, and D. Z. Yu. *Dispersal Ecology*, chapter Oases in the desert : dispersal and host specialization of biotrophic fungal pathogens of plants, pages 395–409. Cambridge Univ. Press, Cambridge, UK, 2002.
- H. Brun, A.-M. Chèvre, B. DL Fitt, S. Powers, A.-L. Besnard, M. Ermel, V. Huteau, B. Marquer, F. Eber, M. Renard, and D. Adrivot. Quantitative resistance increase the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *New Phytol.*, 185 :285–299, 2010.
- J. J. Burdon. The Structure of Pathogen Populations in Natural Plant Communities. *Annu. Rev. Phytopathol.*, 31 :305–23, 1993.
- J. J. Burdon and P. H. Thrall. Pathogen evolution across the agro-ecological interface : implications for disease management. *Evolutionary Applications*, 1(1) :57–65, 2008. doi : 10.1111/j.1752-4571.2007.00005.x.
- J. J. Burdon and P. H. Thrall. What have we learned from studies of wild plant-pathogen associations? —the dynamic interplay of time, space and life-history. *Eur J Plant Pathol*, 138 :417–429, 2014. doi : 10.1007/s10658-013-0265-9.
- S. A. Bustin. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology*, 25 :169–193, 2000.
- A. Caballero. Developments in the prediction of effective population size. *Heredity*, 73 :657–679, 1994.
- V. Caffier, P. Lasserre-Zuber, M. Giraud, M. Lascostes, R. Stievenard, A. Lemarquand, E. van de Weg, P. Expert, C. Denancé, F. Didelot, B. Le Cam, and C.-E. Durel. Erosion of quantitative host resistance in the apple \times *Venturia inaequalis* pathosystem. *Infection, Genetics and Evolution*, 27 :481–489, 2014. doi : <http://dx.doi.org/10.1016/j.meegid.2014.02.003>.
- F. Calenge, A. Faure, M. Goerre, C. Gebhardt, W. E. Van de Weg, L. Parisi, and C.-E. Durel. Quantitative Trait Loci (QTL) Analysis Reveals Both Broad-Spectrum and Isolate-Specific QTL for Scab Resistance in an Apple Progeny Challenged with Eight Isolates of *Venturia inaequalis*. *Phytopathology*, 94 :370–379, 2004.
- C. Caranta, A. Palloix, V. Lefebvre, and A. M. Daubèze. QTLs for a component of partial

- resistance to cucumber mosaic virus in pepper : restriction of virus installation in host-cells. *Theor Appl Genet*, 94 :431–438, 1997.
- P. Carrasco, F. de la Iglesia, and S. F. Elena. Distribution of Fitness and Virulence Effects Caused by Single-Nucleotide Substitutions in *Tobacco Etch Virus*. *J. Virol.*, 81(23) :12979–12984, 2007.
- P. Castagnone-Sereno, M. Bongiovanni, A. Palloix, and A. Dalmaso. Selection for *Meloidogyne incognita* virulence against resistance genes from tomato and pepper and specificity of the virulence/resistance determinants. *Eur J Plant Pathol*, 102 :585–590, 1996.
- N. Champouret, K. Bouwmeester, H. Rietman, T. van der Lee, C. Maliepaard, A. Heupink, P. J. I. Van de Vondervoort, E. Jacobsen, R. G. F. Visser, E. A. G. van der Vossen, F. Govers, and V. G. A. A. Vleeshouwers. *Phytophthora infestans* Isolates Lacking Class I *ipi0* Variants Are Virulent on *Rpi-b1b1* Potato. *MPMI*, 22(12) :1535–1545, 2009.
- L. Chao. Fitness of RNA virus decreased by muller’s ratchet. *Nature*, 348 :454–455, 1990.
- B. Charlesworth. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10 :195–205, 2009.
- J. S. Clark. Why environmental scientists are becoming Bayesians. *Ecology Letters*, 8 :2–14, 2005. doi : 10.1111/j.1461-0248.2004.00702.x.
- J. Colhoun. Ernest Charles Large : pioneer in phytopathometry. *Annu. Rev. Phytopathol.*, 31 : 23–31, 1993.
- J. Corander, P. Waldmann, and M. J. Sillanpaa. Bayesian analysis of genetic differentiation between populations. *Genetics*, 163 :367–374, 2003.
- R. Covas, C. R. Brown, M. D. Anderson, and M. B. Brown. Stabilizing selection on body mass in the sociable weaver *Philetairus socius*. *Proc. R. Soc. Lond. B*, 269 :1905–1909, 2002. doi : 10.1098/rspb.2002.2106.
- C. J. R. Cumagun, R. L. Bowden, J. E. Jurgenson, J. F. Leslie, and T. Miedaner. Genetic mapping of pathogenicity and aggressiveness of *Gibberella zeae* (*Fusarium graminearum*) toward wheat. *Phytopathology*, 94(5) :520–526, 2004.
- C. J. D’Arcy, D. M. Eastburn, and G. L. Schumann. Illustrated glossary of plant pathology. *The Plant Health Instructor*, 2001. doi : 10.1094/PHI-I-2001-0219-01.

- T. Day and S. R. Proulx. A general theory for the evolutionary dynamics of virulence. *The American Naturalist*, 163 :E40–E63, 2004.
- F. de la Iglesia and S. F. Elena. Fitness Declines in *Tobacco Etch Virus* upon Serial Bottleneck Transfer. *J. Virol.*, 81(10) :4941–4947, 2007.
- F. de la Iglesia, F. Martínez, J. Hillung, J. M. Cuevas, P. J. Gerrish, J.-A. Daròs, and S. F. Elena. Luria-delbrück estimation of *Turnip Mosaic Virus* mutation rate *in vivo*. *J. Virol.*, 86(6) :3386–3388, 2012.
- J. de Silva and S. K. Wyatt. Fitness valleys constrain HIV-1's adaptation to its secondary chemokine coreceptor. *J. Evol. Biol.*, 27 :604–615, 2014.
- F. Débarre and S. Gandon. *Introduction à l'épidémiologie intégrative des maladies infectieuses et parasitaires*, chapter Épidémiologie évolutive des maladies infectieuses, pages 347–376. Guéguan, J.-F. and Choisy, M., 2009.
- C.A. Dedryver, A. Le Ralec, and F. Fabre. The conflicting relationships between aphids and men : A review of aphid damage and control strategies. *C. R. Biologies*, 333 :539–553, 2010.
- F. Delmotte, P. Mestre, C. Schneider, H.-H. Kassemeyer, P. Kozma, S. Richart-Cervera, M. Rouxel, and L. Delière. Rapid and multiregional adaptation to host partial resistance in a plant pathogenic oomycete : Evidence from European populations of *Plasmopara viticola*, the causal agent of grapevine downy mildew. *Infection, Genetics and Evolution*, 27 :500–508, 2014. doi : <http://dx.doi.org/10.1016/j.meegid.2013.10.017>.
- J. A. Díaz, C. Nieto, E. Moriones, V. Truniger, and M. A. Aranda. Molecular Characterization of a *Melon necrotic spot virus* Strain That Overcomes the Resistance in Melon and Nonhost Plants. *MPMI*, 17(6) :668–675, 2004.
- K. Dietz. The first epidemic model : a historical note on P.D. En'ko. *Austral. J. Statist.*, 30 (1) :56–65, 1988.
- C. Djian-Caporalino, S. Molinari, A. Palloix, A. Ciancio, A. Fazari, N. Marteu, N. Ris, and P. Castagnone-Sereno. The reproductive potential of the root-knot nematode *Meloidogyne incognita* is affected by selection for virulence against major resistance genes from tomato and pepper. *Eur J Plant Pathol*, 131 :431–440, 2011. doi : [10.1007/s10658-011-9820-4](https://doi.org/10.1007/s10658-011-9820-4).
- C. Djian-Caporalino, A. Palloix, A. Fazari, N. Marteu, A. Barbary, P. Abad, A.-M. Sage-Palloix, T. Mateille, S. Risso, R. Lanza, C. Taussig, and P. Castagnone-Sereno. Pyramiding, alternating

- or mixing : comparative performances of deployment strategies of nematode resistance genes to promote plant resistance efficiency and durability. *BMC Plant Biology*, 14(53) :1–13, 2014.
- D. D. Donalson and R. M. Nibset. Population dynamics and spatial scale : effects of system size on population persistence. *Ecology*, 80 :2492–2507, 1999.
- J. Doumayrou, G. Thébaud, F. Vuillaume, M. Peterschmitt, and C. Urbino. Mapping genetic determinants of viral traits with F_{ST} and quantitative trait locus (QTL) approaches. *Virology*, 484 :346–353, 2015.
- J. W. Drake and J. J. Holland. Mutation rates among RNA viruses. *PNAS*, 96(24) :13910–13913, 1999.
- J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow. Rates of Spontaneous Mutation. *Genetics*, 148 :1667–1686, 1998.
- E. Duarte, D. Clarke, A. Moya, E. Domingo, and J. Holland. Rapid fitness losses in mammalian RNA virus clones due to muller’s ratchet. *Proc. Natl. Acad. Sci. USA*, 89 :6015–6019, 1992.
- H. L. Duhamel de Monceau. Explication physique d’une maladie qui fait périr plusieurs plantes dans le Gastinois, et particulièrement le safran. *Mém. Acad. Sci. Math. Phys. (Paris)*, pages 100–112, 1728.
- B. Dybiec, A. Kleczkowski, and C. A. Gilligan. Controlling disease spread on networks with incomplete knowledge. *Phys. Rev. E*, 70 :066145, 2004.
- S. F. Elena, P. Agudelo-Romero, P. Carrasco, F. M. Codoñer, S. Martín, C. Torres-Barceló, and R. Sanjuán. Experimental evolution of plant RNA viruses. *Heredity*, 100 :478–483, 2008.
- S. F. Elena, S. Bedhomme, P. Carrasco, J. M. Cuevas, F. de la Iglesia, G. Lafforgue, J. Lalić, À. Pròsper, N. Tromas, and M. P. Zwart. The Evolutionary Genetics of Emerging Plant RNA Viruses. *MPMI*, 24(3) :287–293, 2011.
- S. F. Elena, A. Fraile, and F. García-Arenal. *Evolution and Emergence of Plant Viruses*. Burlington : Academic Press, 2014.
- European Environment Agency (EEA). Climate change, impacts and vulnerability in Europe 2012. Technical report, EEA, 2012.
- W. J. Ewens. *Mathematical Population Genetics 1 Theoretical Introduction*. Springer-Verlag, 2004.

- F. Fabre, C. Bruchou, A. Palloix, and B. Moury. Key determinants of resistance durability to plant viruses : Insights from a model linking within- and between-host dynamics. *Virus Research*, 141 :140–149, 2009.
- F. Fabre, J. Montarry, J. Coville, R. Senoussi, V. Simon, and B. Moury. Modelling the evolutionary dynamics of viruses within their hosts : a case study using high-throughput sequencing. *PLoS Pathog*, 8(4) :e1002654, 2012a.
- F. Fabre, E. Rousseau, L. Mailleret, and B. Moury. Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytol.*, 193 :1064–1075, 2012b.
- F. Fabre, E. Rousseau, L. Mailleret, and B. Moury. Epidemiological and evolutionary management of plant resistance : optimizing the deployment of cultivar mixtures in time and space in agricultural landscapes. *Evolutionary Applications*, 8(10) :919–932, 2015.
- L. Fahrig. Effects of habitat loss and fragmentation on population extinction. *Journal of Wildlife Management*, 61(3) :603–610, 1997.
- A. Fereres and B. Raccach. Plant Virus Transmission by Insects. *eLS John Wiley & Sons, Ltd : Chichester*, pages 1–12, 2015. doi : 10.1002/9780470015902.a0000760.pub3.
- R. A. Fisher. On the dominance ratio. *Proc. R. Soc. B Edinburgh*, 52 :321–341, 1922.
- R. A. Fisher. The distribution of gene ratios for rare mutations. *Proc. R. Soc. B Edinburgh*, 50 :205–220, 1930.
- B. D. L. Fitt and H. A. McCartney. *Plant Disease Epidemiology*, volume 1, chapter Spore dispersal in relation to epidemic models, pages 311–345. New York : MacMillan, 1986.
- H. H. Flor. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.*, 9 :275–296, 1971.
- M. Foll, O. E. Gaggiotti, J. T. Daub, A. Vatsiou, and L. Excoffier. Widespread Signals of Convergent Adaptation to High Altitude in Asia and America. *Am. J. Hum. Genet.*, 95 : 394–407, 2014a.
- M. Foll, Y.-P. Poh, N. Renzette, A. Ferrer-Admettla, C. Bank, H. Shim, A.-S. Malaspinas, G. Ewing, P. Liu, D. Wegmann, D. R. Caffrey, K. B. Zeldovich, D. N. Bolon, J. P. Wang, T. F. Kowalik, C. A. Schiffer, R. W. Finberg, and J. D. Jensen. Influenza virus drug resistance : a time-sampled population genetics perspective. *PLoS Genet*, 10(2) :e1004185, 2014b.

- M. Foll, H. Shim, and J. D. Jensen. WFABC : a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources*, 2014c.
- Food and Agriculture Organisation (FAO) of the United Nations. Rome declaration on world food security, November 1996. URL <http://www.fao.org/wfs/>.
- Food and Agriculture Organisation (FAO) of the United Nations. Declaration of the world summit on food security, November 2009.
- S. Fournet, M. C. Kerlan, L. Renault, J. P. Dantec, C. Rouaux, and J. Montarry. Selection of nematodes by resistant plants has implications for local adaptation and cross-virulence. *Plant Pathology*, 62 :184–193, 2013.
- A. Fraile, I. Pagán, G. Anastasio, E. Sáez, and F. García-Arenal. Rapid Genetic Diversification and High Fitness Penalties Associated with Pathogenicity Evolution in a Plant Virus. *Mol. Biol. Evol.*, 28(4) :1425–1437, 2011.
- R. C. French and D. C. Stenger. Evolution of Wheat Streak Mosaic Virus : Dynamics of Population Growth Within Plants May Explain Limited Variation. *Annu. Rev. Phytopathol.*, 41 :199–214, 2003.
- R. Froissart, D. Roze, M. Uzest, L. Galibert, S. Blanc, and Y. Michalakis. Recombination Every Day : Abundant Recombination in a Virus during a Single Multi-Cellular Host Infection. *PLoS Biology*, 3(3) :0389–0395, 2005.
- D. L. Funnell, P. S. Matthews, and H. D. vanEtten. Breeding for Highly Fertile Isolates of *Nectria haematococca* MPVI that are Highly Virulent on Pea and In Planta Selection for Virulent Recombinants. *Phytopathology*, 91 :92–101, 2001.
- S. Gago, S. F. Elena, R. Flores, and R. Sanjuán. Extermely High Mutation Rate of a Hammerhead Viroid. *Science*, 323 :1308, 2009.
- A. Gallais. *Théorie de la sélection en amélioration des plantes*. Masson, Paris, 1990.
- A. P. Galvani. Epidemiology meets evolutionary ecology. *Trends in Ecology and Evolution*, 18 (3) :132–139, 2003.
- S. Gandon and Y. Michalakis. Evolution of parasite virulence against qualitative or quantitative host resistance. *Proc. R. Soc. Lond. B*, 267 :985–990, 2000.

- S. Gandon and Y. Michalakis. Local adaptation, evolutionary potential and host-parasite coevolution : interactions between migration, mutation, population size and generation time. *J. Evol. Biol.*, 15 :451–462, 2002.
- V. V. Ganusov, C. T. Bergstrom, and R. Antia. Within-host population dynamics and the evolution of microparasites in a heterogeneous host population. *Evolution*, 56(2) :213–223, 2002.
- F. García-Arenal and B. A. McDonald. An Analysis of the Durability of Resistance to Plant Viruses. *Phytopathology*, 93 :941–952, 2003.
- F. García-Arenal, A. Fraile, and J. M. Malpica. Variability and Genetic Structure of Plant Virus Populations. *Annu. Rev. Phytopathol.*, 39 :157–86, 2001.
- F. García-Arenal, A. Fraile, and J. M. Malpica. Variation and evolution of plant virus populations. *Int Microbiol*, 6 :225–232, 2003.
- T. Garnett, M. C. Appleby, A. Balmford, I. J. Bateman, T. G. Benton, P. Bloomer, B. Burlingame, M. Dawkins, L. Dolan, D. Fraser, M. Herrero, I. Hoffmann, P. Smith, P. K. Thornton, C. Toulmin, S. J. Vermeulen, and H. C. J. Godfray. Sustainable Intensification in Agriculture : Premises and Policies. *Science*, 341 :33–34, 2013.
- K. A. Garrett, G. A. Forbes, S. Savary, P. Skelsey, A. H. Sparks, C. Valdivia, A. H. C. van Bruggen, L. Willocquet, A. Djurle, E. Duveiller, H. Eckersten, S. Pande, C. Vera Cruz, and J. Yuen. Complexity in climate-change impacts : an analytical framework for effects mediated by plant disease. *Plant Pathology*, 60 :15–30, 2011.
- P. J. Gerrish and R. E. Lenski. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102/103 :127–144, 1998.
- D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25) :2340–2361, 1977.
- C. A. Gilligan. Sustainable agriculture and plant diseases : an epidemiological perspective. *Philos. Trans. R. Soc.*, 363 :741–759, 2008.
- H. C. J. Godfray, J. R. Beddington, I. R. Crute, L. Haddad, D. Lawrence, J. F. Muir, J. Pretty, S. Robinson, S. M. Thomas, and C. Toulmin. Food Security : The Challenge of Feeding 9 Billion People. *Science*, 327 :812–818, 2010.

- P. Gómez, A. M. Rodríguez-Hernández, B. Moury, and M. A. Aranda. Genetic resistance for the sustainable control of plant virus to diseases : breeding, mechanisms and durability. *Eur J Plant Pathol*, 125(1) :1–22, 2009.
- P. González-Jara, A. Fraile, T. Canto, and F. García-Arenal. The Multiplicity of Infection of a Plant Virus Varies during Colonization of Its Eukaryotic Host. *J. Virol.*, 83(15) :7487–7494, 2009.
- P. González-Jara, A. Moreno-Letelier, A. Fraile, D. Piñero, and F. García-Arenal. Impact of Human Management of the Genetic Variation of Wild Pepper, *Capsicum annuum* var. *glabriusculum*. *PLoS ONE*, 6(12) :e28715, 2011.
- L. Gout, M. L. Kuhn, L. Vincenot, S. Bernard-Samain, L. Cattolico, M. Barbetti, O. Moreno-Rico, M.-H. Balesdent, and T. Rouxel. Genome structure impacts molecular evolution at the *AvrLm1* avirulence locus of the plant pathogen *Leptosphaeria maculans*. *Environmental Microbiology*, 9(12) :2978–2992, 2007.
- S. Gubbins and C. A. Gilligan. Biological control in a disturbed environment. *Phil. Trans. R. Soc. Lond. B*, 352 :1935–1949, 1997a.
- S. Gubbins and C. A. Gilligan. Persistence of Host-parasite Interactions in a Disturbed Environment. *J. theor. Biol.*, 188 :241–258, 1997b.
- J. Gustavsson, C. Cederberg, U. Sonesson, R. van Otterdijk, and A. Meybeck. Global Food Losses and Food Waste : Extent, causes and Prevention. Technical report, FAO, Rome, Italy, 2011.
- S. Gutiérrez, M. Yvon, G. Thébaud, B. Monsion, Y. Michalakis, and S. Blanc. Dynamics of the Multiplicity of Cellular Infection in a Plant Virus. *PLoS Pathog*, 6(9) :e1001113, 2010.
- S. Gutiérrez, Y. Michalakis, and S. Blanc. Virus population bottlenecks during within-host progression and host-to-host transmission. *Current Opinion in Virology*, 2 :546–555, 2012a.
- S. Gutiérrez, M. Yvon, E. Pirolles, E. Garzo, A. Fereres, Y. Michalakis, and S. Blanc. Circulating Virus Load Determines the Size of Bottlenecks in Viral Populations Progressing within a Host. *PLoS Pathog*, 8(11) :e1003009, 2012b.
- S. Gutiérrez, E. Pirolles, M. Yvon, V. Baecker, Y. Michalakis, and S. Blanc. The multiplicity of cellular infection changes depending on the route of cell infection in a plant virus. *J Virol*, 89 :9665–9675, 2015. doi : doi:10.1128/JVI.00537-15.

- K. E. Hammond-Kosack and J. E. Parker. Deciphering plant-pathogen communication : fresh perspectives for molecular resistance breeding. *Current Opinion in Biotechnology*, 14 :177–193, 2003.
- B. D. Harrison. Virus variation in relation to resistance-breaking in plants. *Euphytica*, 124 : 181–192, 2002.
- J. M. Heffernan and L. M. Wahl. The effects of genetic drift in experimental evolution. *Theoretical Population Biology*, 62 :349–356, 2002.
- J. Hermisson and P. S. Pennings. Soft sweeps : molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169 :2335–2352, 2005.
- F. Holmes. Inheritance of resistance to tobacco mosaic disease in tobacco. *Phytopathology*, 28 : 553–61, 1938.
- C. J. R. Illingworth, L. Parts, S. Schiffels, G. Liti, and V. Mustonen. Quantifying selection acting on a complex trait using allele frequency time series data. *Mol. Biol. Evol.*, 29(4) : 1187–1197, 2012. doi : 10.1093/molbev/msr289.
- L. A. Imhof and M. A. Nowak. Evolutionary game Dynamics in a Wright-Fisher process. *J. Math. Biol.*, 52 :667–681, 2006.
- Y. Iwasa, F. Michor, and M. A. Nowak. Stochastic tunnels in evolutionary dynamics. *Genetics*, 166 :1571–1579, 2004.
- W. C. James. Assessment of plant diseases and losses. *Annu. Rev. Phytopathol.*, 12 :27–48, 1974.
- B. Janzac, F. Fabre, A. Palloix, and B. Moury. Constraints on evolution of virus avirulence factors predict the durability of corresponding plant resistances. *Molecular Plant Pathology*, 10(5) :599–610, 2009.
- B. Janzac, J. Montarry, A. Palloix, O. Navaud, and B. Moury. A point mutation in the polymerase of potato virus Y confers virulence toward the *Pvr4* resistance of pepper and a high competitiveness cost in susceptible cultivar. *MPMI*, 23 (6) :823–830, 2010.
- M.J. Jeger. Analysis of Disease Progress as a Basis for Evaluating Disease Management Practices. *Annu. Rev. Phytopathol.*, 42 :61–82, 2004.

- C. E. Jenner, F. Sánchez, S. B. Nettleship, G. D. Foster, F. Ponz, and J. A. Walsh. The Cylindrical Inclusion Gene of *Turnip mosaic virus* Encodes a PaPathogen Determinant to the Brassica Resistance Gene *TuRB01*. *MPMI*, 13(10) :1102–1108, 2000.
- N. F. Jensen. Intra-varietal diversification in oat breeding. *Agron. J.*, 44 :50–54, 1952.
- R. Johnson. A critical analysis of durable resistance. *Annu. Rev. Phytopathol.*, 22 :309–30, 1984.
- T. Johnson. Man-Guided Evolution in Plant Rusts. *Science*, 133(3450) :357–362, 1961.
- P. E. Jorde and N. Ryman. Unbiased estimator for genetic drift and effective population size. *Genetics*, 177 :927–935, 2007.
- B.-C. Kang, I. Yeam, and M. M. Jahn. Genetics of Plant Virus Resistance. *Annu. Rev. Phytopathol.*, 45 :581–621, 2005.
- S. Kang, M. H. Lebrun, L. Farrall, and B. Valent. Gain of Virulence Caused by Insertion of a Pot3 Transposon in a *Magnaporthe grisea* Avirulence Gene. *MPMI*, 14(5) :671–674, 2001.
- S. Karlin. Rates of Approach to Homozygosity for Finite Stochastic Models with Variable Population Size. *The American Naturalist*, 102(927) :443–455, 1968.
- R. Kassen and T. Bataillon. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nature Genetics*, 38(4) :484–488, 2006. doi : 10.1038/ng1751.
- B. Kearney, P. C. Ronald, D. Dahlbeck, and B. J. Staskawicz. Molecular basis for evasion of plant host defence in bacterial spot disease of pepper. *Nature*, 332(7) :541–543, 1988.
- F. Keesing, L. K. Belden, P. Daszak, A. Dobson, C. D. Harvell, R. D. Holt, P. Hudson, A. Jolles, K. E. Jones, C. E. Mitchell, S. S. Myers, T. Bogish, and R. S. Ostfeld. Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, 468 :647–652, 2010.
- W.O. Kermack and A.G. McKendrick. Contributions to the mathematical-theory of epidemics. *Proceedings of the Royal Society*, 115A :700–721, 1927.
- H. B. D. Kettlewell. Recognition of Appropriate Backgrounds by the Pale and Black Phases of Lepidoptera. *Nature*, 175 :943–944, 1955.

- M. Khelifa, D. Massé, S. Blanc, and M. Drucker. Evaluation of the minimal time of *Cauliflower mosaic virus* in different hosts. *Virology*, 396 :238–245, 2010.
- H. J. Kim, J.-H. Han, J. H. Yoo, H. J. Cho, and B.-D. Kim. Development of a sequence characteristic amplified region marker linked to the L^4 locus conferring broad spectrum resistance to tobamoviruses in pepper plants. *Mol. Cells.*, 25(2) :205–210, 2008.
- M. Kimura. Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harbor Symposia on Quantitative Biology*, 20 :33–53, 1955.
- M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47 : 713–719, 1962.
- M. Kimura. *Mathematical Topics in Population Genetics*, volume 1. Springer-Verlag, 1970.
- M. Kimura and J. F. Crow. The measurement of effective population number. *Evolution*, 17(3) : 279–288, 1963.
- T. Kitakado, S. Kitada, H. Kishino, and H. J. Skaug. An integrated-likelihood method for estimating genetic differentiation between populations. *Genetics*, 173 :2073–2082, 2006.
- S. Kiyosawa. Genetics and epidemiological modeling of breakdown of plant disease resistance. *Annu. Rev. Phytopathol.*, 20 :93–117, 1982.
- S. Koizumi. Rice blast control with multilines in japan. *Mew TW, Borrmeo E, Hardy B. Exploiting Biodiversity for Sustainable Pest Management. Manila : International Rice Research Institute*, pages 143–157, 2001.
- C. S. Kousik and D. F. Ritchie. Race shift in *Xanthomonas campestris* pv. *vesicatoria* within a season in field-grown pepper. *Phytopathology*, 86 :952–958, 1996.
- J. G. Kühn. *Die Krankheiten der Kukturgewächse*. Bosselman, Berlin, 1858.
- M. Lacerda and C. Seoighe. Population genetics inference for longitudinally-sampled mutants under strong selection. *Genetics*, 199(3) :1237–1250, 2014. doi : 10.1534/genetics.114.167957.
- D. Laehnemann, A. Borkhardt, and A. C. McHardy. Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. *Briefings in Bioinformatics*, 17(1) : 154–179, 2016. doi : 10.1093/bib/bbv029.

- G. Lafforgue, N. Tromas, S. F. Elena, and M. P. Zwart. Dynamics of the establishment of systemic *Potyvirus* infection : Independent yet cumulative action of primary infection sites. *J. Virol.*, 86(23) :12912–12822, 2012.
- A.-L. Laine, J. J. Burdon, P. N. Dodds, and P. H. Thrall. Spatial variation in disease resistance : from molecules to metapopulations. *Journal of Ecology*, 99 :96–112, 2011. doi : 10.1111/j.1365-2745.2010.01738.x.
- J. Lalić and S. F. Elena. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. *Heredity*, 109 :71–77, 2012.
- R. Lande and G. F. Barrowclough. *Viable populations for conservation*, chapter Effective population size, genetic variation and their use in population management, pages 87–123. Cambridge Univ. Press, New York, 1987.
- R. Lanfear, H. Kokko, and A. Eyre-Walker. Population size and the rate of evolution. *Trends in Ecology and Evolution*, 29(1) :33–41, 2014. doi : 10.1016/j.tree.2013.09.009.
- C. Lannou. Variation and selection of quantitative traits in plant pathogens. *Annu. Rev. Phytopathol.*, 50 :319–38, 2012.
- C. Lannou, C. de Vallavieille-Pope, C. Biass, and H. Goyeau. The efficacy of mixtures of susceptible and resistant hosts to two wheat rusts of different lesion size : controlled condition experiments and computerized simulations. *J. Phytopathology*, 140 :227–237, 1994.
- E. C. Large. The interpretation of progress curves for potato late blight and other plant diseases. *Plant Pathology*, 1 :109–117, 1952.
- E. Lázaro, C. Escarmís, J. Pérez-Mercader, S. C. Manrubia, and E. Domingo. Resistance of virus to extinction on bottleneck passages : Study of a decaying and fluctuating pattern of fitness loss. *PNAS*, 100(19) :10830–10835, 2003.
- A. Lê Van, V. Caffier, P. Lasserre-Zuber, A. Chauveau, D. Brunel, B. Le Cam, and C.-E. Durel. Differential selection pressures exerted by host resistance quantitative trait loci on a pathogen population : a case study in apple \times *Venturia inaequalis* pathosystem. *New Phytol.*, 197 : 899–908, 2013. doi : 10.1111/nph.12086.
- J. E. Leach, C. M. Vera Cruz, J. Bai, and H. Leung. Pathogen Fitness Penalty as a Predictor of Durability of Disease Resistance Genes. *Annu. Rev. Phytopathol.*, 39 :187–224, 2001.

- H. Lecoq, B. Moury, C. Desbiez, A. Palloix, and M. Pitrat. Durable virus resistance in plants through conventional approaches : a challenge. *Virus Research*, 100 :31–39, 2004.
- T. Lenormand. Gene flow and the limits to natural selection. *TRENDS in Ecology and Evolution*, 17(4) :183–189, 2002.
- K. J. Leonard. Selection pressures and plant pathogens. *Annals of the New York Academy of Sciences*, 287(1) :207–222, 1977.
- R. Levins and M. Wilson. Ecological theory and pest management. *Ann. Rev. Entomol.*, 25 : 287–308, 1980.
- M. Z. Levy, A. Tustin, R. Castillo-Neyra, T. S. Mabud, K. Levy, C. M. Barbu, V. R. Quispe-Machaca, J. Ancca-Juarez, K. Borrini-Mayori, C. Naquira-Velarde, R. S. Ostfeld, and the Chagas Working Group in Arequipa. Bottlenecks in domestic animal populations can facilitate the emergence of *Trypanosoma cruzi*, the aetiological agent of Chagas disease. *Proc. R. Soc. B*, 282 :1–9, 2015.
- Y. Li, T. A. J. van der Lee, A. Evenhuis, G. B. M. van den Bosch, P. J. van Bekkum, M. G. Förch, M. P. E. van Gent-Pelzer, H. M. G. van Raaij, E. Jacobsen, S. W. Huang, F. Govers, V. G. A. A. Vleeshouwers, and G. J. T. Kessel. Population Dynamics of *Phytophthora infestans* in the Netherlands Reveals Expansion and Spread of Dominant Clonal Lineages and Virulence in Sexual Offspring. *Genes Genomes Genet.*, 2 :1529–1540, 2012.
- J. F. Limbrunner, R. M. Vogel, L. C. Brown, and M. ASCE. Estimation of harmonic mean of a lognormal variable. *Journal of hydrologic engineering*, 1(59) :59–66, 2000.
- M. Lind, K. Dalman, J. Stenlid, B. Karlsson, and Å Olson. Identification of quantitative trait loci affecting virulence in the basidiomycete *Heterobasidion annosum* s.l. *Curr Genet*, 52 : 35–44, 2007. doi : 10.1007/s00294-007-0137-y.
- G. Lo Iacono, F. van den Bosch, and C. Gilligan. Durable resistance to crop pathogens : an epidemiological framework to predict risk under uncertainty. *PLoS Comput Biol*, 9(1) : e1002870, 2013.
- G. Luikart, J.-M. Cornuet, and F. W. Allendorf. Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conservation Biology*, 13 (3) :523–530, 1999.

- L. V. Madden and F. van den Bosch. A Population-Dynamics Approach to Assess the Threat of Plant Pathogens as Biological Weapons against Annual Crops. *BioScience*, 52(1) :65–74, 2002.
- L. V. Madden, G. Hughes, and F. van den Bosch. *The Study of Plant Disease Epidemics*. APS Press, 2007.
- T. Magoç and S. L. Salzberg. FLASH : fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21) :2957–2963, 2011.
- L. Mailleret and F. Grognaard. Optimal release policy for prophylactic biological control. *Positive Systems*, pages 89–96, 2006.
- L. Mailleret and F. Grognaard. Global stability and optimisation of a general impulsive biological control model. *Mathematical Biosciences*, 221 :91–100, 2009. doi : 10.1016/j.mbs.2009.07.002.
- L. Mailleret and V. Lemesle. A note on semi-discrete modelling in the life sciences. *Philosophical Transactions of the Royal Society*, 367 :4979–4999, 2009.
- L. Mailleret, M. Castel, J. Montarry, and F. M. Hamelin. From elaborate to compact seasonal plant epidemic models and back ; is competitive exclusion in the details ? *Theoretical ecology*, 5(3) :311–324, 2012.
- A.-S. Malaspinas. Methods to characterize selective sweeps using time serial samples : an ancient DNA perspective. *Molecular Ecology*, 25 :24–41, 2016. doi : 10.1111/mec.13492.
- A.-S. Malaspinas, O. Malaspinas, S. N. Evans, and M. Slatkin. Estimating allele age and selection coefficient from time-serial data. *Genetics*, 192 :599–607, 2012.
- J. M. Malpica, A. Fraile, I. Moreno, C. I. Obies, J. W. Drake, and F. García-Arenal. The Rate and Character of Spontaneous Mutation in an RNA Virus. *Genetics*, 162 :1505–1511, 2002.
- T. C. Marcel, B. Gorguet, M. T. Ta, Z. Kohutova, A. Vels, and R. E. Niks. Isolate specificity of quantitative trait loci for partial resistance of barley to *Puccinia hordei* confirmed in mapping populations and near-isogenic lines. *New Phytol.*, 177 :743–755, 2008.
- F. Martínez, J. Sardanyés, S. F. Elena, and J.-A. Daròs. Dynamics of a Plant RNA Virus Intracellular Accumulation : Stamping Machine *vs.* Geometric Replication. *Genetics*, 188 : 637–646, 2011.

- I. Mathieson and G. McVean. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics*, 193 :973–984, 2013.
- R. M. May and R. M. Anderson. Epidemiology and genetics in the coevolution of parasites and hosts. *Proc. R. Soc. Lond. B*, 219 :281–313, 1983.
- R. M. May and M. A. Nowak. Superinfection, metapopulation dynamics, and the evolution of diversity. *J. Theor. Biol.*, 170 :95–114, 1994.
- B. A. McDonald and C. Linde. Pathogen Population Genetics, Evolutionary Potential, and Durable Resistance. *Annu. Rev. Phytopathol.*, 40 :349–79, 2002a.
- B. A. McDonald and C. Linde. The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica*, 124 :163–180, 2002b.
- W. H. McNeill. *Plagues and Peoples*. Doubleday, London, 1976.
- N. Mideo, S. Alizon, and T. Day. Linking within- and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends in Ecology and Evolution*, 23(9) :511–517, 2008.
- M. Miras, R. N. Sempere, J. J. Kraft, W. A. Miller, M. A. Aranda, and V. Truniger. Interfamilial recombination between viruses led to acquisition of a novel translation-enhancing RNA element that allows resistance breaking. *New Phytologist*, 202 :233–246, 2014.
- S. Miyashita and H. Kishino. Estimation of the size of genetic bottlenecks in cell-to-cell movement of *Soil-Borne Wheat Mosaic Virus* and the possible role of the bottlenecks in speeding up selection of variations in *trans*-acting genes of elements. *J. Virol.*, 84(4) :1828–1837, 2010.
- B. Monsion, R. Froissart, Y. Michalakakis, and S. Blanc. Large bottleneck size in *Cauliflower Mosaic Virus* populations during host plant colonization. *PLoS Pathog*, 4(10) :e1000174, 2008.
- J. Montarry, R. Corbière, and D. Andrivon. Is there a trade-off between aggressiveness and overwinter survival in *Phytophthora infestans*. *Functional Ecology*, 21 :603–610, 2007. doi : 10.1111/j.1365-2435.2007.01252.x.
- J. Montarry, J. Doumayrou, V. Simon, and B. Moury. Genetic background matters : a plant-virus gene-for-gene interaction is strongly influenced by genetic contexts. *Mol. Plant Pathol.*, 12(9) :911–920, 2011.

- J. Montarry, E. Cartier, M. Jacquemond, A. Palloix, and B. Moury. Virus adaptation to quantitative plant resistance : erosion or breakdown? *J. Evol. Biol.*, 25(11) :2242–2252, 2012.
- A. D. Morgan, S. Gandon, and A. Buckling. The effect of migration on local adaptation in a coevolving host-parasite system. *Nature*, 437(8) :253–256, 2005. doi : 10.1038/nature03913.
- B. Moss. Water pollution by agriculture. *Phil. Trans. R. Soc. B*, 363 :659–666, 2008.
- U. Motro and G. Thomson. On heterozygosity and the effective size of populations subject to size changes. *Evolution*, 36(5) :1059–1066, 1982.
- B. Moury, A. Palloix, K. Gebre Selassie, and G. Marchoux. Hypersensitive resistance in tomato spotted wilt virus in three *Capsicum chinense* accessions is controlled by a single gene and is overcome by virulent strains. *Euphytica*, 94 :45–52, 1997.
- B. Moury, C. Morel, E. Johansen, L. Guilbaud, S. Souche, V. Ayme, C. Caranta, A. Palloix, and M. Jacquemond. Mutations in *Potato virus Y* Genome-Linked Protein Determine Virulence Toward Recessive Resistances in *Capsicum annuum* and *Lycopersicon hirsutum*. *MPMI*, 17 (3) :322–329, 2004.
- B. Moury, F. Fabre, and R. Senoussi. Estimation of the number of virus particles transmitted by an insect vector. *PNAS*, 104 :17891–17896, 2007.
- B. Moury, F. Fabre, J. Montarry, B. Janzac, V. Ayme, and A. Palloix. L'adaptation des virus de plantes aux résistances variétales. *Virologie*, 14 :227–239, 2010.
- B. Moury, B. Caromel, E. Johansen, V. Simon, L. Chauvin, E. Jacquot, C. Kerlan, and V. Lefebvre. The Helper Component Proteinase Cistron of *Potato virus Y* Induces Hypersensitivity and Resistance in Potato Genotypes Carrying Dominant Resistance Genes on Chromosome IV. *MPMI*, 24(7) :787–797, 2011a. doi : 10.1094/MPMI-10-10-0246.
- B. Moury, A. Fereres, F. García-Arenal, and H. Lecoq. *Recent Advances in Plant Virology*, chapter Sustainable Management of Plant Resistance to Viruses, pages 219–236. Caister Academic Press, Norwich, UK, 2011b.
- B. Moury, B. Janzac, Y. Ruellan, V. Simon, M. B. Khalifa, H. Fakhfakh, F. Fabre, and A. Palloix. Interaction patterns between *Potato Virus Y* and eif4e-mediated recessive resistance in the *Solanaceae*. *Journal of Virology*, 88(17) :9799–9807, 2014. doi : 10.1128/JVI.00930-14.

- N. Movahedi, G. Paillotin, and A. Neveu. Le monde manquera-t-il de terres pour nourrir les hommes du XXIème siècle? *Comptes rendus de l'Académie d'agriculture de France*, 95(3) : 75–80, 2009.
- H. J. Muller. Some genetic aspects of sex. *Am. Nat.*, 66 :118–138, 1932. doi : 10.1086/280418.
- H. J. Muller. The relation of recombination to mutational advance. *Mutat. Res.*, 106 :2–9, 1964.
- C. C. Mundt. Use of Multiline Cultivars and Cultivar Mixtures for Disease Management. *Annu. Rev. Phytopathol.*, 40 :381–410, 2002.
- C. C. Mundt. Durable resistance : A key to sustainable management of pathogens and pests. *Infection, Genetics and Evolution*, 27 :446–455, 2014. doi : <http://dx.doi.org/10.1016/j.meegid.2014.01.011>.
- C. C. Mundt and J. A. Browning. Development of Crown Rust Epidemics in Genetically Diverse Oat Populations : Effect of Genotype Unit Area. *Phytopathology*, 75(5) :607–610, 1985.
- J.D. Murray. *Mathematical Biology : I. An Introduction, Third Edition*. Springer-Verlag, 2002.
- L. R. Nault. Arthropod Transmission of Plant Viruses : A New Synthesis. *Ann. Entomol. Soc. Am.*, 90 :521–541, 1997.
- M. Nei and F. Tajima. Genetic drift and estimation of effective population size. *Genetics*, 98 : 625–640, 1981.
- A. Ohtsuki and A. Sasaki. Epidemiology and disease-control under gene-for-gene plant-pathogen interaction. *Journal of Theoretical Biology*, 238 :780–794, 2006.
- R. S. Ostfeld and F. Keesing. Effects of Host Diversity on Infectious Disease. *Annu. Rev. Phytopathol.*, 43 :157–82, 2012.
- S. P. Otto and M. C. Whitlock. The Probability of Fixation in Populations of Changing Size. *Genetics*, 146 :723–733, 1997.
- I. Pagán, P. González-Jara, A. Moreno-Letelier, M. Rodelo-Urrego, A. Fraile, D. Piñero, and F. García-Arenal. Effect of Biodiversity Changes in Disease Risk : Exploring Disease Emergence in a Plant-Virus System. *PLoS Pathog*, 8(7) :e1002796, 2012.

- A. Palloix, V. Ayme, and B. Moury. Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. *New Phytol.*, 183 :190–199, 2009.
- J. C. Panetta. A mathematical model of periodically pulsed chemotherapy : tumor recurrence and metastasis in a competitive environment. *Bulletin of Mathematical Biology*, 58(3) : 425–447, 1996.
- J. C. Panetta. A Mathematical Model of Drug Resistance : Heterogeneous Tumors. *Mathematical Biosciences*, 147 :41–61, 1998.
- J. Papaïx. Structure du paysage agricole et risque épidémique, une approche démo-génétique. Master’s thesis, AgroParisTech, 2011.
- J. Papaïx, O. David, C. Lannou, and H. Monod. Dynamics of Adaptation in Spatially Heterogeneous Metapopulations. *PLoS ONE*, 8(2) :e54697, 2013. doi : 10.1371/journal.pone.0054697.
- J. Papaïx, S. Touzeau, H. Monod, and C. Lannou. Can epidemic control be achieved by altering landscape connectivity in agricultural systems? *Ecological Modelling*, 284 :35–47, 2014.
- J. Papaïx, J. J. Burdon, J. Zhan, and P. H. Thrall. Crop pathogen emergence and evolution in agro-ecological landscapes. *Evolutionary Applications*, 8 :385–402, 2015.
- B. Pariaud, V. Ravigné, F. Halkett, H. Goyeau, J. Carlier, and C. Lannou. Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathology*, 58 :409–424, 2009.
- M. Park, C. Loverdo, S. J. Schreiber, and J. O. Lloyd-Smith. Multiple scales of selection influence the evolutionary emergence of novel pathogens. *Phil. Trans. R. Soc. B*, 368 : 20120333, 2013.
- J. E. Parlevliet. Durability of resistance against fungal, bacterial and viral pathogens ; present situation. *Euphytica*, 124 :147–156, 2002.
- Z. Patwa and L. M. Wahl. The fixation probability of beneficial mutations. *J. R. Soc. Interface*, 5 :1279–1289, 2008.
- W. L. Pedersen and S. Leath. Pyramiding major genes for resistance to maintain residual effects. *Annu. Rev. Phytopathol.*, 26 :369–78, 1988.

- O. Pierrugues, L. Guilbaud, I. Fernandez-Delmond, F. Fabre, M. Tepfer, and M. Jacquemond. Biological properties and relative fitness of inter-subgroup cucumber mosaic virus RNA 3 recombinants produced *in vivo*. *Journal of General Virology*, 88 :2852–2861, 2007. doi : 10.1099/vir.0.83077-0.
- D. Pink and I. Puddephat. Deployment of disease resistance genes by plant transformation – a 'mix and match' approach. *Trends in plant science*, 4(2) :71–75, 1999.
- D. A. C. Pink. Strategies using genes for non-durable disease resistance. *Euphytica*, 124 : 227–236, 2002.
- F. J. Poelwijk, S. Tănase-Nicola, D. J. Kiviet, and S. J. Tans. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology*, 272 : 141–144, 2011.
- S. G. Potts, J. C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger, and W. E. Kunin. Global pollinator declines : trends, impacts and drivers. *Trends in Ecology and Evolution*, 25(6) : 345–353, 2010.
- A. G. Power. Plant community diversity, herbivore movement, and an insect-transmitted disease of maize. *Ecology*, 68(6) :1658–1669, 1987. doi : 10.2307/1939858.
- A. G. Power. Virus spread and vector dynamics in genegenetic diverse plant populations. *Ecology*, 72 :232–41, 1991.
- J. Quenouille. *Bases génétiques et fonctionnelles de la durabilité des résistances polygéniques au virus Y de la pomme de terre (PVY) chez le piment (Capsicum annuum)*. PhD thesis, Université d'Avignon et des Pays de Vaucluse, Février 2013.
- J. Quenouille, J. Montarry, A. Palloix, and B. Moury. Farther, slower, stronger : how the plant genetic background protects a major resistance gene from breakdown. *Molecular Plant Pathology*, 14 :109–118, 2013.
- J. Quenouille, E. Paulhiac, B. Moury, and A. Palloix. Quantitative trait loci from the host genetic background modulate the durability of a resistance gene : a rational basis for sustainable resistance breeding plants. *Heredity*, 112 :579–587, 2014.
- J. Quenouille, Saint-Felix L., B. Moury, and A. Palloix. Diversity of genetic backgrounds modulating the durability of a major resistance gene. analysis of a core collection of pepper landraces resistant to *Potato virus Y*. *Molecular Plant Pathology*, 17(2) :296–302, 2016.

- R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL <http://www.R-project.org/>.
- R. R. Regoes, M. A. Nowak, and S. Bonhoeffer. Evolution of virulence in a heterogeneous host population. *Evolution*, 54(1) :64–71, 2000.
- O. Restif and J. C. Koella. Concurrent evolution of resistance and tolerance to pathogens. *The American Naturalist*, 164(4) :E90–E102, 2004.
- REX Consortium. Heterogeneity of selection and the evolution of resistance. *Trends in Ecology and Evolution*, 28(2) :110–118, 2013.
- T. D. Rhodes, O. Nikolaitchik, J. Chen, D. Powell, and W.-S. Hu. Genetic Recombination of Human Immunodeficiency Virus Type 1 in One Round of Viral Replication : Effects of Genetic Distance, Target Cells, Accessory Genes, and Lack of High Negative Interference in Crossover Events. *J. Virol.*, 79(3) :1666–1677, 2005.
- C. Richards and P. L. Leberg. Temporal changes in allele frequencies and a population's history of severe bottlenecks. *Conservation Biology*, 10 (3) :832–839, 1996.
- C. Riedel, A. Habekuß, E. Schliephake, R. Niks, I. Broer, and F. Ordon. Pyramiding of *Ryd2* and *Ryd3* conferring tolerance to a german isolate of *Barley yellow dwarf virus-pav* (bydv-pav-asl-1) leads to quantitative resistance against this isolate. *Theor Appl Genet*, 123 :69–76, 2011. doi : 10.1007/s00122-011-1567-y.
- C. Robaglia and C. Caranta. Translation initiation factors : a weak link in plant RNA virus infection. *Trends in Plant Science*, 11(1) :40–45, 2006.
- G. Rodrigo, M. P. Zwart, and S. F. Elena. Onset of virus systemic infection in plants is determined by speed of cell-to-cell movement and number of primary infection foci. *J. R Soc. Interface*, 11 :20140555, 2014.
- E. Rousseau, B. Moury, L. Mailleret, R. Senoussi, A. Palloix, V. Simon, S. Valière, F. Grognard, and F. Fabre. Joint estimation of effective population size and selection coefficient without neutral markers : Method validation and application to the experimental evolution of viruses. *PLoS Genetics*, Submitted.
- T. Rouxel and M.-H. Balesdent. *Encyclopedia of Life Sciences (ELS)*, chapter Avirulence Genes, pages 1–13. John Wiley & Sons, Ltd, Chichester, 2010. doi : 10.1002/9780470015902.a0021267.

- T. Rouxel, A. Penaud, X. Pinochet, H. Brun, L. Gout, R. Delourme, J. Schmit, and M.-H. Balesdent. A 10-year survey of populations of *Leptosphaeria maculans* in France indicates a rapid adaptation towards the *Rlm1* resistance gene of oilseed rape. *European Journal of Plant Pathology*, 109 :871–881, 2003.
- I. M. Rouzine, A. Rodrigo, and J. M. Coffin. Transition between stochastic evolution and deterministic evolution in the presence of selection : general theory and application in virology. *Microbiol. Mol. Biol. Rev.*, 65 :151–185, 2001.
- L. Råberg, D. Sim, and A. F. Read. Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science*, 318 :812–814, 2007.
- S. Sacristán and F. García-Arenal. The evolution of virulence and pathogenicity in plant pathogen populations. *Mol. Plant Pathol.*, 9(3) :369–384, 2008.
- S. Sacristán, J. M. Malpica, A. Fraile, and F. García-Arenal. Estimation of Population Bottlenecks during Systemic Movement of *Tobacco Mosaic Virus* in Tobacco Plants. *J. Virol.*, 77(18) :9906–9911, 2003.
- M. Salathé, A. Scherer, and S. Bonhoeffer. Neutral drift and polymorphism in gene-for-gene systems. *Ecology Letters*, 8(9) :925–932, 2005. doi : 10.1111/j.1461-0248.2005.00794.x.
- R. Sanjuán. Mutational fitness effects in RNA and single-stranded DNA viruses : common patterns revealed by site-directed mutagenesis studies. *Phil. Trans. R. Soc. B*, 365 :1975–1982, 2010.
- R. Sanjuán, A. Moya, and S. F. Elena. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *PNAS*, 101(22) :8396–8401, 2004.
- S. Schürch, C. C. Linde, W. Knogge, L. F. Jackson, and B. A. McDonald. Molecular Population Genetic Analysis Differentiates Two Avirulence Mechanisms of the Fungal Avirulence Gene *NIP1*. *MPMI*, 17(10) :1114–1125, 2004.
- M. W. Shaw. Seasonally induced chaotic dynamics and their implications in models of plant disease. *Plant Pathol.*, 43(5) :790–801, 1994.
- J. S. Shonkwiler and N. Hanley. A New Approach to Random Utility Modeling using the Dirichlet Multinomial Distribution. *Environmental and Resource Economics*, 26 :401–416, 2003.

- P. D. Sniegowski and P. J. Gerrish. Beneficial mutations and the dynamics of adaptation in asexual populations. *Phil. Trans. R. Soc. B*, 365 :1255–1263, 2010.
- K. Soetaert, T. Petzoldt, and R. W. Setzer. Solving Differential Equations in R : Package deSolve. *Journal of Statistical Software*, 33(9) :1–25, 2010. URL <http://www.jstatsoft.org/v33/i09>.
- R.V. Sole, R. Ferrer, I. Gonzalez-Garcia, J. Quer, and E. Domingo. Red queen dynamics, competition and critical points in a model of RNA virus quasispecies. *Journal of Theoretical Biology*, 198 :47–59, 1999.
- J. Song, J. M. Bradeen, S. K. Naess, J. A. Raasch, S. M. Wielgus, G. T. Haberlach, J. Liu, H. Kuang, S. Austin-Phillips, C. R. Buell, J. P. Helgeson, and J. Jiang. Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. *PNAS*, 100 (16) :9128–9133, 2003.
- M. Steinrücken, A. Bhaskar, and Y. S. Song. A novel spectral method for inferring general diploid selection from time series genetic data. *Ann Appl Stat.*, 8(4) :2203–2222, 2014.
- E. H. Stukenbrock and B. A. McDonald. The origins of plant pathogens in agro-ecosystems. *Annu. Rev. Phytopathol.*, 46 :75–100, 2008. doi : 10.1146/annurev.phyto.010708.154114.
- F. Suffert, I. Sache, and C. Lannou. Early stages of septoria tritici blotch epidemics of winter wheat : build-up, overseasoning, and release of primary inoculum. *Plant Pathology*, 60 : 166–177, 2011.
- L. Tamisier, E. Rousseau, S. Baraillé, G. Nemouchi, M. Szadkowski, L. Mailleret, F. Grogard, F. Fabre, B. Moury, and A. Palloix. Quantitative trait loci in pepper genome control the effective population size of two RNA viruses at inoculation. In *XVI Eucarpia Capsicum and Eggplant Meeting*, Kecskemét, Hungary, 2016.
- M. Y. A. Tan, R. Alles, R. C. B. Hutten, R. G. F. Visser, and H. J. van Eck. Pyramiding of *Meloidogyne hapla* resistance genes in potato does not result in an increase of resistance. *Potato Research*, 52 :331–340, 2009.
- M. Y. A. Tan, R. C. B. Hutten, R. G. F. Visser, and H. J. van Eck. The effect of pyramiding *Phytophthora infestans* resistance genes $R_{Pi-mcd1}$ and R_{Pi-ber} in potato. *Theor Appl Genet*, 121 :117–125, 2010.

- P. Tataru, T. Bataillon, and A. Hobolt. Inference under a wright-fisher model using an accurate beta approximation. *Genetics*, 201 :1133–1141, 2015. doi : 10.1534/genetics.115.179606.
- P. S. Teng and R. E. Gaunt. Modelling systems of disease and yield loss in cereals. *Agricultural Systems*, 6(2) :131–154, 1980.
- J. Terhorst, C. Schlötterer, and Y. S. Song. Multi-locus analysis of genomic time series data from experimental evolution. *PLoS Genet*, 11(4) :e1005069, 2015. doi : 10.1371/journal.pgen.1005069.
- M. Tester and P. Langridge. Breeding Technologies to Increase Crop Production in a Changing World. *Science*, 327 :818–822, 2010.
- J. N. Thompson and J. J. Burdon. Gene-for-gene coevolution between plants and parasites. *Nature*, 360 :121–125, 1992.
- P. H. Thrall and J. J. Burdon. The spatial scale of pathogen dispersal : Consequences for disease dynamics and persistence. *Evolutionary Ecology Research*, 1 :681–701, 1999.
- D. Tilman. Global environmental impacts of agricultural expansion : The need for sustainable and efficient practices. *Proc. Natl. Acad. Sci. USA*, 96 :5995–6000, 1999.
- N. Tromas and S. F. Elena. The Rate and Spectrum of Spontaneous Mutations in a Plant RNA Virus. *Genetics*, 185 :983–989, 2010.
- N. Tromas, M. P. Zwart, M. Poulain, and S. F. Elena. Estimation of the *in vivo* recombination rate for a plant RNA virus. *Journal of General Virology*, 95 :724–732, 2014. doi : 10.1099/vir.0.060822-0.
- J. E. Truscott, C. R. Webb, and C. A. Gilligan. Asymptotic analysis of an epidemic model with primary and secondary infection. *Bulletin of Mathematical Biology*, 59(6) :1101–1123, 1997.
- T. Tscharntke, Y. Clough, T. C. Wanger, L. Jackson, I. Motzke, I. Perfecto, J. Vandermeer, and A. Whitbread. Global food security, biodiversity conservation and the future of agricultural intensification. *Biological Conservation*, 151 :53–59, 2012.
- T. L. Turner, A. D. Stewart, A. T. Fields, W. R. Rice, and A. M. Tarone. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genet*, 7(3) :e1001336, 2011. doi : 10.1371/journal.pgen.1001336.

- C. Vacher, D. Bourguet, F. Rousset, C. Chevillon, and M. Hochberg. Modelling the spatial configuration of refuges for a sustainable control of pests : a case study of *Bt* cotton. *J. Evol. Biol.*, 16 :378–387, 2003.
- M. van Baalen. Coevolution of recovery ability and virulence. *Proc. R. Soc. Lond. B*, 265 : 317–325, 1998.
- F. van den Berg, N. Bacaer, J. A. J. Metz, C. Lannou, and F. van den Bosch. Periodic host absence can select for higher or lower parasite transmission rates. *Evol Ecol*, 25 :121–137, 2011.
- F. van den Berg, C. Lannou, C. Gilligan, and F. van den Bosch. Quantitative resistance can lead to evolutionary changes in traits not targeted by the resistance QTLs. *Evolutionary Applications*, 7 :370–380, 2014.
- F. van den Bosch and C.A. Gilligan. Measures of durability of resistance. *Phytopathology*, 93 : 616–625, 2003.
- J. E. van der Plank. *Plant Pathology : An Advanced Treatise, Vol. 3 : The Disease Population*, volume 3, chapter Analysis of epidemics, pages 229–289. Academic Press, New York, 1960.
- J. E. van der Plank. *Plant Diseases : Epidemics and Control*. Academic Press, New York, 1963.
- J. E. van der Plank. *Disease Resistance in Plants*. Academic Press, Inc., New York, London, 1968.
- E. van der Vossen, A. Sikkema, B. te Lintel Hekkert, J. Gros, P. Stevens, M. Muskens, D. Wouters, A. Pereira, W. Stiekema, and S. Allefs. An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. *The Plant Journal*, 36 :867–882, 2003.
- R. Vitalis, M. Gautier, K. J. Dawson, and M. A. Beaumont. Detecting and Measuring Selection from Gene Frequency Data. *Genetics*, 196 :799–817, 2014.
- J. A. Vucetich, T. A. Waite, and L. Nunney. Fluctuating population size and the ratio of effective to census population size. *Evolution*, 51(6) :2017–2021, 1997.
- L. M. Wahl and P. J. Gerrish. The probability that beneficial mutations are lost in populations with periodic bottleneck. *Evolution*, 55(12) :2606–2610, 2001.

- L. M. Wahl, P. J. Gerrish, and I. Saika-Voivod. Evaluating the Impact of Population Bottlenecks in Experimental Evolution. *Genetics*, 162 :961–971, 2002.
- G.-L. Wang, D. J. Mackill, J. M. Bonman, S. R. McCouch, M. C. Champoux, and R. J. Nelson. RFLP Mapping of Genes Conferring Complete and Partial Resistance to Blast in a Durably Resistant Rice Cultivar. *Genetics*, 136 :1421–1434, 1994.
- R. S. Waples. A Generalized Approach for Estimating Effective Population Size From Temporal Changes in Allele Frequency. *Genetics*, 121 :379–391, 1989.
- R. S. Waples, T. Antao, and G. Luikart. Effects of Overlapping Generations on Linkage Disequilibrium Estimates of Effective Population Size. *Genetics*, 197 :769–780, 2014.
- M. C. Whitlock and R. Gomulkiewicz. Probability of fixation in a heterogeneous environment. *Genetics*, 171 :1407–1417, 2005.
- D. J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman & Hall, 2006.
- E. G. Williamson and M. Slatkin. Using Maximum Likelihood to Estimate Population Size From Temporal Changes in Allele Frequencies. *Genetics*, 152 :755–761, 1999.
- M. S. Wolfe. The Current Status and Prospects of Multiline Cultivars and Variety Mixtures for Disease Resistance. *Annu. Rev. Phytopathol.*, 23 :251–73, 1985.
- S. Wright. Evolution in Mendelian populations. *Genetics*, 16(2) :97–159, 1931.
- S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proceedings of the Sixth International Congress of Genetics*, pages 356–366, 1932.
- S. Wright. Statistical genetics in relation to evolution. In Hermann, editor, *Exposés de Biométrie et de Statistique Biologique*, Paris, 1939.
- J. C. Zadoks and L. M. Koster. A historical survey of botanical epidemiology : A sketch of the development of ideas in ecological phytopathology. *Mededelingen Landbouwhogeschool Wageningen (The Netherlands)*, 76(12), 1976.
- F. Zanini and R. A. Neher. FFPopSim : an efficient forward simulation package for the evolution of large populations. *Bioinformatics*, 28 (24) :3332–3333, 2012a.
- F. Zanini and R. A. Neher. *FFPopSim : An efficient forward simulation package for the evolution of large populations*. Max Planck Institute for Developmental Biology, September 2012b.

- J. Zhan, P. H. Thrall, and J. J. Burdon. Achieving sustainable plant disease management through evolutionary principles. *Trends in Plant Science*, 19(9) :570–575, 2014.
- J. Zhan, P. H. Thrall, J. Papaïx, L. Xie, and J. J. Burdon. Playing on a pathogen’s weakness : Using evolution to guide sustainable plant disease control strategies. *Annu. Rev. Phytopathol.*, 53 :2.1–2.25, 2015.
- X.-S. Zhang and J. Holt. Mathematical Models of Cross Protection in the Epidemiology of Plant-Virus Diseases. *Phytopathology*, 91 :924–934, 2001.
- M. P. Zwart and S. F. Elena. Matters of size : Genetic bottlenecks in virus infection and their potential impact on evolution. *Annu. Rev. Phytopathol.*, 2 :161–79, 2015.
- M. P. Zwart, J.-A. Daròs, and S. F. Elena. One is enough : *in vivo* effective population size is dose-dependent for a plant RNA virus. *PLoS Pathog*, 7(7) :e1002122, 2011.

Publications

- Accepted
- L. Tamisier, E. Rousseau, S. Baraillé, G. Nemouchi, M. Szadkowski, L. Mailleret, F. Grogard, F. Fabre, B. Moury and A. Palloix (2016) Quantitative trait loci in pepper genome control the effective population size of two RNA viruses at inoculation. *XVI Eucarpia Capsicum and Eggplant Meeting* Kecskemét, Hungary.
 - F. Fabre, E. Rousseau, L. Mailleret and B. Moury (2015) Epidemiological and evolutionary management of plant resistance: optimizing the deployment of cultivar mixtures in time and space in agricultural landscapes. *Evolutionary Applications* 8(10):919–932.
 - F. Fabre, E. Rousseau, L. Mailleret and B. Moury (2012) Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytologist* 193: 1064–1075.
- Submitted
- E. Rousseau, B. Moury, L. Mailleret, R. Senoussi, A. Palloix, V. Simon, S. Valière, F. Grogard and F. Fabre. Joint estimation of effective population size and selection coefficient without neutral markers: method validation and application to the experimental evolution of viruses. Submitted to *PLoS Genetics*.
- In preparation
- E. Rousseau, M. Bonneault, F. Fabre, B. Moury, L. Mailleret and F. Grogard. Modulating genetic drift exerted on virus populations to increase the durability of plant resistance.
 - E. Rousseau, L. Tamisier, F. Fabre, V. Simon, M. Szadkowski, G. Girardot, L. Mailleret, F. Grogard, A. Palloix and B. Moury. Impact of genetic drift, selection and accumulation level on virus adaptation to their host plants.

Talks

- 06/2014 **ECMTB (European Conference on Mathematical and Theoretical Biology)**, *Göteborg (Sweden)*.
Protection of the durability of major resistance genes to plant viruses through quantitative resistance, a modeling approach.
- 01/2013 **RVV ('Rencontres de Virologie Végétale', plant virology meeting)**, *Aussois (France)*.
Sustainable plant resistance management in agricultural landscapes.
- 12/2012 **Workshop Pollution Modeling and Control**, *La Rochelle (France)*.
Durabilité des stratégies de gestion de plantes résistantes aux virus dans un paysage agricole.
- 05/2012 **M3D (Mathematics and Decision for Sustainable Development) meeting**, *Oléron (France)*.
Durabilité des stratégies de déploiement de plantes résistantes dans un paysage agricole.

Posters

- 05/2015 **SMBE (Society for Molecular Biology & Evolution) meeting 'Investigating biological adaptation with NGS: data and models'**, *Hameau de l'étoile (France)*.
Joint estimation of the strength of genetic drift and selection from Next Generation Sequencing time-sampled data: a case study on the adaptation of virus populations to host plant resistance.
- 04/2014 **Jacques Monod Conference 'From emerging to pandemic viruses: interplay between host ecology and viral evolution'**, *Roscoff (France)*.
Protecting the durability of major resistance genes to plant viruses with quantitative resistance.
- 06/2013 **Workshop 'Bridging Theoretical and Experimental Evolution'**, *La Fouly (Switzerland)*.
Impact of quantitative plant resistance on within-host viral demo-genetic dynamics.
- 06/2013 **INRA SPE department Ph.D. meeting**, *Montpellier (France)*.
Impact of quantitative plant resistance on within-host viral demo-genetic dynamics.

Effet de la dérive génétique et de la sélection sur la durabilité de la résistance des plantes aux virus.

Une plante peut être totalement protégée d'un agent pathogène grâce à un gène majeur de résistance, mais ce dernier peut être rapidement contourné suite à l'apparition et à la propagation de variants pathogènes adaptés. Cette thèse s'intéresse aux mécanismes évolutifs permettant le ralentissement de ce contournement chez les virus de plantes en agissant sur deux forces évolutives majeures, la dérive génétique et la sélection, depuis le niveau de l'hôte jusqu'à celui de la parcelle. D'abord, un modèle épidémiologique stochastique de type SI au niveau d'une parcelle agricole a montré que la dérive génétique pouvait être particulièrement bénéfique au rendement agricole lorsque l'adaptation du virus au gène majeur induit un coût de fitness intermédiaire dans les plantes sensibles. Ensuite, la conception et la validation d'un modèle basé sur des équations déterministes de Lotka-Volterra et des processus stochastiques Dirichlet-multinomiaux a permis de distinguer les effets de la dérive génétique et ceux de la sélection sur des données temporelles de compétition intra-plante entre variants viraux, et de mettre en évidence le contrôle génétique de ces effets par les plantes. Enfin, une analyse de la corrélation entre ces estimations des intensités de dérive génétique et de sélection et une estimation expérimentale de la durabilité d'un gène majeur a montré qu'une forte dérive génétique lors des stades précoces de l'infection augmentait la durabilité du gène majeur. Ces résultats ouvrent des perspectives pour une gestion plus durable de la résistance des plantes, par la sélection de variétés de plantes induisant une forte dérive génétique sur les populations d'agents pathogènes.

Mots clés : dérive génétique, sélection, durabilité des résistances, évolution expérimentale, épidémiologie.

Effect of genetic drift and selection on plant resistance durability to viruses.

Plants can be fully protected from their pathogens when they carry major resistance genes, but the efficiency of these genes is limited by the emergence and spread of adapted, resistance-breaking pathogen variants. This thesis studies how evolutionary forces imposed by the plants on pathogen populations may increase the durability of major resistance genes. Using plant viruses as a biological model, this thesis investigates the effect of genetic drift and selection, from the within-host to the host population level. Firstly, a stochastic epidemiological SI model at the field level showed that genetic drift could be particularly beneficial for crop yield when the fitness cost associated with virus adaptation to resistance was intermediate in susceptible plants. Then, the design and validation of a mechanistic-statistical model based on deterministic Lotka-Volterra equations and stochastic Dirichlet-multinomial processes allowed to disentangle the effects of genetic drift from those of selection on temporal data of within-host competition between virus variants. The intensities of genetic drift and selection acting on virus populations were shown to be controlled genetically by the hosts. Finally, a correlation analysis between these estimations of genetic drift and selection intensities and an experimental estimation of the durability of a major resistance gene showed that strong genetic drift during the early stages of plant infection led to an increase in resistance durability. These results open new perspectives for more durable management of plant resistance, by breeding plant varieties inducing strong genetic drift on pathogen populations.

Keywords : genetic drift, selection, resistance durability, experimental evolution, epidemiology.