



HAL
open science

Genomic and ecological bases of aphids adaptation in the context of fruit tree domestication

Sergio Gabriel Olvera Vázquez

► **To cite this version:**

Sergio Gabriel Olvera Vázquez. Genomic and ecological bases of aphids adaptation in the context of fruit tree domestication. Populations and Evolution [q-bio.PE]. Université Paris-Saclay, 2023. English. NNT : 2023UPASB063 . tel-04639088

HAL Id: tel-04639088

<https://theses.hal.science/tel-04639088>

Submitted on 8 Jul 2024

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.

Genomic and ecological bases of aphids adaptation in the context of fruit tree domestication

*Bases génomiques et écologiques de l'adaptation des pucerons dans le
contexte de la domestication des arbres fruitiers*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 567, Sciences du végétal : du gène à l'écosystème (SEVE)
Spécialité de doctorat : Evolution
Graduate School : Biosphera. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **GQE-Le Moulon** (Université Paris-Saclay,
INRAE, CNRS, AgroParisTech), sous la direction d'**Amandine CORNILLE** Chargée de
recherche

Thèse soutenue à Paris-Saclay, le 06 novembre 2023, par

Sergio Gabriel OLVERA VÁZQUEZ

Composition du Jury

Membres du jury avec voix délibérative

Yvan RHABÉ

Directeur de recherche, INRAE Président
Université de Lyon

Jean PECCOUD

Maître de conférences (HDR), Rapporteur & Examinateur
Université de Poitiers

Carole SMADJA

Directrice de recherche, CNRS Rapporteuse & Examinatrice
Université de Montpellier

Jeanne ROPARS

Chargée de recherche, CNRS Examinatrice
Université Paris-Saclay

Titre : Bases génomiques et écologiques de l'adaptation des pucerons dans le contexte de la domestication des arbres fruitiers

Mots clés : puceron, chimioréception, jardin commun, divergence, interaction hôte-parasite, adaptation locale

Résumé : Les contributions relatives des facteurs biotiques et abiotiques à l'adaptation locale restent un sujet de recherche activement étudié. Comprendre les processus d'adaptation locale dans les interactions hôtes-parasites peut aider à résoudre des problèmes urgents, telle que la manière dont les changements environnementaux modifient l'émergence de parasites, conduisant à l'extinction de l'hôte. Ma thèse vise à évaluer les bases génomiques et écologiques de l'adaptation des pucerons, et est divisée en quatre chapitres. En particulier, nous avons étudié la présence de patrons d'adaptation (locale) aux niveaux phénotypiques et génomiques en lien avec la colonisation récente du pommier cultivé (*Malus domestica*) par *Dysaphis plantaginea* en Europe. Dans le chapitre 1, nous avons étudié les gènes impliqués dans l'adaptation des pucerons à leur hôte à une échelle macroévolutive. Nous avons montré un taux d'évolution plus élevé dans une sous-famille de pucerons connue pour infester de nombreux hôtes, entraîné par l'expansion et la contraction des gènes qui ont des fonctions liées à l'adaptation des pucerons à leurs hôtes. Nous avons montré une faible synténie des gènes des récepteurs olfactifs (OR) et gustatifs (GR), évoluant en répétitions en tandem selon le modèle naissance-mort et sous sélection purifiante. Les éléments transposables (ET) n'ont joué aucun rôle dans l'évolution des gènes GR et OR. Cependant, nous avons trouvé une dynamique d'ET spécifique à chaque espèce, avec une activité des ET très récente chez les espèces de pucerons infestant les arbres fruitiers. Dans le chapitre 2, je me suis intéressé aux processus d'adaptation à une échelle microévolutive. De manière préliminaire, j'ai reconstitué l'histoire de la colonisation de *D. plantaginea* sur le pommier cultivé à l'aide de marqueurs microsatellites. Nous avons détecté cinq populations aux États-Unis, Maroc, Espagne de l'Est et de l'Ouest. Une faible différenciation génétique entre populations et une diversité génétique élevée est observée, sauf au Maroc et en Amérique du Nord, probablement en raison de récents événements de colonisation dans ces régions. Des simulations basées sur la coalescence a également indiqué de forts flux de gènes entre les populations. Dans le chapitre 3, j'ai utilisé des polymorphismes de nucléotidique unique (SNPs) pour étudier les bases génomiques de l'adaptation (locale) chez *D. plantaginea* et ses deux espèces sœurs, *Dysaphis devectora* et *Dysaphis pyri*. Les trois espèces forment des groupes génétiques distincts, et *D. plantaginea* se divise en quatre populations distinctes en Amérique du Nord, Israël, Maroc/Espagne et dans les pays européens. Nous présentons les résultats préliminaires des scans des génomes dans les six groupes génétiques, qui révèlent des SNP candidats en lien avec l'adaptation du puceron à son environnement, mais des investigations plus approfondies sont nécessaires. Dans le chapitre 4, j'étudie les bases écologiques de l'adaptation locale de *D. plantaginea* en utilisant une approche expérimentale sur le terrain. Nos résultats préliminaires suggèrent une maladaptation locale des populations de pucerons français et une adaptation globale des populations espagnoles. Une analyse plus approfondie est nécessaire pour clarifier ces patrons d'adaptation. Les résultats de ce doctorat apportent des informations pour mieux comprendre comment les conditions biotiques (hôte) et abiotiques peuvent façonner l'architecture génomique des insectes phytophages, conduisant à une adaptation (locale). Les résultats constituent des bases de données et des connaissances sans précédent pour les futurs programmes de lutte contre les ravageurs.

Title: Genomic and ecological bases of aphids adaptation in the context of fruit tree domestication

Keywords: aphid, chemoreception, common garden, divergence, host-parasite interaction, local adaptation

Abstract: The relative contributions of biotic and abiotic factors to local adaptation still need to be studied. Understanding local adaptation processes in host-parasite interactions will also help to tackle pressing issues, such as how environmental changes alter the emergence of parasites, leading to host extinction. My Ph.D. project consisted of four main chapters to evaluate the genomic and ecological bases of adaptation in aphids. In particular, we investigated whether (local) adaptation occurred during the recent colonization of cultivated apple (*Malus domestica*) by *Dysaphis plantaginea* in Europe and what are the genomic bases associated with it. We investigated different scientific hypotheses at different evolutionary scales. In Chapter 1, we studied genes involved in aphid adaptation to their host at a macroevolutionary scale. We showed a higher evolutionary rate in an aphid subfamily known for its large host breadth and including significant crop pests, driven by gene expansion and contraction with functions related to aphid adaptation to their hosts. We showed low synteny of olfactory receptors (OR) and gustatory receptors (GR) genes, evolving in tandem repeats following the birth-death model and under purifying selection. Transposable elements (TEs) did not have a role in the GR and OR gene evolution. However, we found species-specific TE dynamics, with major TEs burst in aphid infesting fruit trees. In Chapter 2, I used a microevolutionary lens and reconstructed the colonization history of *D. plantaginea* on the cultivated apple using microsatellite markers. We detected five populations in the US, Morocco, Western and Eastern Europe, and Spain. Populations showed weak genetic differentiation and high genetic diversity, except in Morocco and North America, likely due to recent colonization events there. Machine learning combined with coalescent-based simulations also indicated substantial gene flow among populations. In Chapter 3, I used single nucleotide polymorphisms (SNPs) to investigate genomic patterns of (local) adaptation in *D. plantaginea* and its two sister species, *Dysaphis devectora* and *Dysaphis pyri*. The three species formed distinct genetic groups, and *D. plantaginea* showed four distinct populations in North America, Israel, Morocco/Spain, and the rest of the European countries. I present preliminary results of genome scans in the six genetic groups detected before, which revealed candidate outlier SNPs for adaptation, but further investigations are needed. In Chapter 4, I study the ecological bases of *D. plantaginea* local adaptation using an experimental approach. Our preliminary results suggest local maladaptation of the French aphid population while global adaptation of the Spanish population. Further analysis is needed to clarify these patterns of adaptation. The results of this Ph.D. project will bring information to understand how abiotic and abiotic conditions can shape genomic architecture in phytophagous insects, leading to (local) adaptation. We also provide unprecedented databases and knowledge for future pest management programs.

Acknowledgments

I would first thank my supervisor, Amandine Cornille, who is my thesis supervisor. She has guided me patiently and encouraged me through her passion for developing projects and finding the best way to decipher the project's scientific questions. I was lucky to learn how to be a promising future supervisor from Amandine. She spent time elucidating each doubt I had, advising and listening to me when I faced difficulties during my Ph.D. project and facing challenges during life in France. I am also grateful to have the opportunity to observe how Amandine was settling the ECLECTIC laboratory, recruiting each of our esteemed colleagues, and the way she has been leading the team. I also appreciate how she encouraged me and facilitated the resources to attend workshops, courses, congress, and symposia. These scientific events allow me to learn new concepts, improve my knowledge, and acquire skills to analyze my research project, promote my project, and start my work networking.

I want to thank each member of my Ph.D. committee, Jacqui SHYKOFF, Tatiana GIRAUD, Élodie MARCHADIER, and Camille ROUX, whose gave me their time every year to give me remarks, feedback, and confront my knowledge throughout their critical questions based on their high expertise in the different areas treated in my Ph.D. project. They also allowed me to discuss with them and request some advice on topics unrelated to my research project. I also appreciate every comment and discussion based on their professional and life experience because it enriches my professional development.

I am also highly thankful to Xilong CHEN, who was highly patient in teaching, reviewing, and explaining different methodologies and bioinformatics pipelines. He guided me since my arrival on the team in the new process of acquiring all the bioinformatic skills but also gave some of his time to have informal discussions about our lives. I also would like to thank the permanent members of the ECLECTIC and GQE-Le Moulon group, Carine REMOUÉ, Anthony VENON, and Amandine CORNILLE. They have been fundamental elements in the performance of my Ph.D. project, teaching,

guiding, and supporting me in the different methodologies in the laboratory and experimental field. I was super lucky to learn a lot from three qualified experts in the laboratory and field methodologies and how to deal with daily life in France. I also would like to thank each of the current temporal members of the ECLECTIC team, Célia LOUGMANI, Yuqi NIE, Ronan DODOLE, Maxime CRIADO, and Aurélie MESNIL (citing them in the way they arrived at the lab), each of them has spent some of their time to discuss my result, teaching some methodologies, and giving some of their time to enjoy the life outside the lab. I also would like to thank the temporal researcher, Yonghua JIANG, who supported me with a challenging and key part of my thesis project, the aphid rearing. Also, Shuo LIU has been giving me some advice and taking time for informal discussions.

I also appreciate the support of the Director of the GQE-Le Mounon UMR, Christine DILLMAN, and the administrative area of the GQE-Le Moulon, Valérie LESPINAS, Sandrine LE-BIHAN, Élodie CHALARD, and Jeff ALAN ASSI. I appreciate all your work and time processing all the information and guiding me in the paperwork. They were fundamental in having all the elements to carry on my organization of a workshop, experiments, having the material, supporting me in attending the scientific conferences, and going to perform my experiment in Spain.

I want to thank the bioinformatic and communication area of the GQE-Le Moulon Benoît JOHANNET, Adrien FALCE, and Olivier LANGELLA for supporting me in every issue that I have with my computer, cluster, and the IDEEV, GQE-Le Moulon website.

At the beginning of my arrival at the laboratory, I had the opportunity to meet another team full of expert researchers, the GEVAD. Each team member gave me some of their time attending my presentations, giving me some advice, and enriching my knowledge through their exciting questions. In particular, Karine ALIX taught me a lot about a new, challenging, and fascinating area: the Transposable Elements. She also supported me in developing the section related to TEs in my Ph.D. project. She also gave me her precious time in the training process to compete in the doctoral school, and I

appreciate the time she took to go to my workspace to ask how I was. I also thank Maud FAGNY, who supported me upon my arrival in the lab and guided me in the primary methodologies of my Ph.D. project. I highly appreciate the support of the technical part of the team, who taught and supported me in the experiments' rules and performance in the laboratory at the beginning of my Ph.D.

I thank all the expert researchers from the research teams involved in my Ph.D. project. They were supporting, guiding, and helping me to reach and improve my results. I thank all interns involved in the aphid cross-infestation experiment performance. In particular, Cristina DE LOS REYES RAMOS, Ramón GONZALEZ, Imane ELFAR, and Lilian BROUILLARD allowed me experience by supervising their work in some sections of my Ph.D. project.

I am also grateful to the researchers and technicians of the research teams that comprised the GQE – Le Moulon UMR. They have been supporting me in giving advice, essential questions, and suggesting procedures to face challenges during the execution of my Ph.D. project. In addition, I thank you for all the good moments that each person at the IDEEV shared with me during the event participation.

I would like to thank the University of Paris-Saclay, which founded my Ph.D. project for four years and awarded me a grant for my trip to Spain to perform the aphid cross-infestation experiment. I also thank the ATIP-Inserm-CNRS program, which found my Ph.D. project.

I would like to thank my family, particularly Mike ANAYA-SAINZ, who supported me during my Ph.D. project. He helped me perform the aphid cross-infestation experiment (preparing the material and execution of the infestation), sample leaves and aphids, and create figures. He encouraged me to face the challenges during the different stages of my Ph.D. project.

Preface

This thesis dissertation aims to bring key information about the genomic and ecological bases of the adaptation of the rosy apple aphid to its environment. I performed this PhD thesis work at the Génétique Quantitative et Évolution (GQE-Le Moulon) at the *Institut Diversité, Écologie et Évolution* (IDEEV) under the supervision of Dr. Amandine CORNILLE, CNRS researcher and leader of the *Ecologie et Génomique des interactions multi-espèces* (ECLECTIC) group. The *Université Paris-Saclay* funded my Ph.D. contract for three years, and I was awarded an additional year because of the COVID-19 lockdown, which has impacted my Ph.D. project, and one month because of the moving of the *GQE-Le Moulon* at the *IDEEV* in 2022. I would like to thank the University of Paris-Saclay, which founded my Ph.D. project for four years and awarded me a grant for my trip to Spain to perform the aphid cross-infestation experiment. I also thank the ATIP-Inserm-CNRS program, which found my Ph.D. project. This work represents an international collaboration.

We investigated the genomic and ecological bases of the rosy apple aphid local adaptation using comparative genomics, population genetics, population genomics, and experimental approaches. My Ph.D. project consisted of four main chapters. First, I studied genes involved in the adaptation of aphids to their hosts at a large evolutionary scale (Chapter 1). Second, I reconstructed the colonization history of *D. plantaginea* to its obligate host, the cultivated apple (Chapter 2). Third, I investigated genomic patterns of local adaptation in *D. plantaginea* (Chapter 3). Finally, I studied the ecological bases of *D. plantaginea* local adaptation using an experimental approach (Chapter 4). A summary of each of these chapters is presented below.

In Chapter 1, we investigated the evolutionary processes at play during the evolution of aphids at a large evolutionary scale. We wanted to provide insight into the evolutionary processes shaping gene molecular evolution in phytophagous insects. This chapter will be soon submitted to the journal *Molecular Biology and Evolution*.

We reconstructed the colonization history of the rosy apple aphid and published

our findings (Chapter 2) in the Peers Community in Evolutionary Biology (<https://doi.org/10.24072/pcjournal.26>). Will also divulgate our results in an article in the online journal "Passion Entomologie" (<https://passion-entomologie.fr/pommier-puceron-cendre/>).

We are investigating the genomics and ecological bases of (local) adaptation in the rosy apple aphid. We are presenting our preliminary results in Chapters 3 and 4. The original findings of this thesis dissertation will add to our understanding of how biotic and abiotic conditions can shape genomic architecture in aphids, ultimately leading to (local) adaptation.

Sergio Gabriel OLVERA VAZQUEZ

Université Paris-Saclay

1. CONTENT

1. Content	9
2. Résumé général	11
3. Introduction	41
3.1 Annexes and extra activities.....	60
3.2 Student supervision.....	60
3.3 References.....	62
4. Outline Chapter 1 : Chemosensory genes and transposable elements evolutionary dynamics in aphids	76
4.1 Chapter 1 : Chemosensory genes and transposable elements evolutionary dynamics in aphids (manuscript under submission).	90
4.1.1 .TE family divergence.....	122
4.2 Text S1. DNA extraction and <i>Dysaphis plantaginea</i> genome sequencing and assembly	149
4.3 Supplementary material.....	154
5. Outline Chapter 2 : Colonization history of the rosy apple aphid	275
5.1 Large-scale geography survey provides insights into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa (first published manuscript)	281
6. Outline chapter 3 : Genomic bases of local adaptation in the rosy apple aphid (<i>Dysaphis plantaginea</i>) to its host, the cultivated apple (<i>Malus domestica</i>)	282
6.1 Genomic bases of local adaptation in the rosy apple aphid (<i>Dysaphis plantaginea</i>) to its host, the cultivated apple (<i>Malus domestica</i>) (Manuscript in preparation).....	289
6.2 Introduction	290
6.3 Material and methods	295
6.4 Results.....	300
6.5 Discussion and Perspectives	308
6.6 Conclusion	310
6.8 References.....	311
6.9 Supplementary text.....	320
6.10 Supplementary Material	323
7. Outline chapter 4 : An experimental test for local adaptation of the rosy apple aphid (<i>Dysaphis plantaginea</i>) during its recent rapid colonization on its cultivated apple host (<i>Malus domestica</i>) in Europe	352
7.1 An experimental test for local adaptation of the rosy apple aphid (<i>Dysaphis plantaginea</i>) during its recent rapid colonization on its cultivated apple host (<i>Malus domestica</i>) in Europe (Manuscript in	

preparation).....	358
7.2 Introduction	360
7.3 Material and Methods.....	362
7.4 Preliminary results.....	371
7.5 Discussion and Perspectives	373
7.6 References	376
7.7 Supplementary text.....	380
7.8 Supplementary Material	386
7.9 Annexes: Experimental test for local adaptation of the rosy apple aphid (<i>Dysaphis plantaginea</i>) during its recent paid colonization on its cultivated apple host (<i>Malus domestica</i>) in Europe	392
8. General discussion and conclusion	393
8.1 General conclusion.....	403
8.2 References	405

2. RESUME GENERAL

Les biologistes de l'évolution tentent de comprendre les processus évolutifs qui ont façonné des populations ou des espèces génétiquement différenciées en réponse à l'environnement (Wolf et Ellegren, 2017). Parmi ces processus, la divergence prend place lorsque des populations originellement semblables se séparent géographiquement ou écologiquement, suite à des pressions sélectives variées. Cette séparation progressive entraîne une différenciation génétique au sein des populations, se traduisant par des fréquences alléliques différentes et des caractéristiques phénotypiques uniques. La divergence peut résulter de processus démographiques (dérive ou/et flux génétique) ou adaptatifs (sélection).

Le contexte de la domestication est particulièrement pertinent pour l'étude de la divergence des parasites, qui présentent de fréquents changements d'hôtes (Gladieux et al., 2015 ; Powell et al., 2013 ; Stukenbrock et McDonald, 2008). Les changements d'hôte impliquent une colonisation rapide du parasite vers de nouvelles régions géographiques. Il est alors essentiel de savoir si la colonisation rapide de différents environnements par les parasites, suite à des changements d'hôtes, est le résultat au moins en partie, de divergence adaptative locale. De plus, quelles sont les bases génomiques associées à l'adaptation locale dans ce contexte ? Plus généralement, quelle est l'architecture génomique sous-jacente (nombre, localisation et effets des régions génomiques) à la divergence adaptative (locale) ? Quels sont les processus génomiques qui permettent ces changements adaptatifs (substitution d'acides aminés, duplications de gènes, gains de gènes, pertes de gènes, sélection positive et purificatrice impliquée) ? Comment les facteurs abiotiques ou biotiques contribuent-ils à la divergence adaptative (locale) ? Répondre à ces questions est fondamental pour élucider les mécanismes d'adaptation des populations à leur environnement (Weigel & Nordborg, 2015). Au cours de mon doctorat, j'ai étudié ces questions en

combinant génomique comparative, génomique des populations et approches expérimentales, en utilisant le puceron cendré du pommier (*Dysaphis plantaginea*), un ravageur majeur de la pomme cultivée, comme système biologique.

L'adaptation locale représente la phase initiale de l'isolement reproductif et de la spéciation écologique (Lenormand, 2012 ; Karvonen et Seehausen, 2012), et son étude permet d'évaluer la contribution de facteurs environnementaux contribuant à la divergence adaptative (Savolainen et al., 2013). L'adaptation locale est le résultat de la variation environnementale spatiale de la sélection, conduisant à des génotypes adaptés localement avec valeur adaptative (taux de survie ou de performance) plus élevée dans leur environnement local que dans les environnements abiotiques et/ou biotiques étrangers (Kawecki et Ebert, 2004 ; Savolainen et al., 2007 ; Wadgymar et al., 2022). Les jardins communs sont des outils permettant de détecter les patrons d'adaptation locale, et ils ont été utilisés principalement chez plantes dans un contexte abiotique (Leimu et Fischer, 2008 ; Kawecki et Ebert, 2004 ; Blanquart et al., 2013 ; Sinclair et al., 2015 ; Nooten et Hughes, 2017 ; Capblancq et al., 2023 ; Meek et al., 2023 ; Razaque et al., 2023). Pourtant, chez les parasites des plantes, une adaptation locale du parasite peut théoriquement être attendue. Les parasites, ayant, en général, un taux d'évolution plus élevé que leur hôte, ils peuvent ainsi présenter des patrons d'adaptation locale à travers la distribution de l'hôte (facteurs biotiques). Mais cela peut aussi dépendre des environnements locaux (Gandon et al., 2008 ; Brown et Tellier, 2011). D'autres outils intéressants à utiliser sont des expériences de transplantations réciproques multi-populations dans des environnements biotiques et abiotiques reflétant l'hétérogénéité spatiale dans lequel le parasite est retrouvé. Or, ce type d'expériences, nécessaire pour comprendre le rôle de l'environnement biotique et abiotique dans l'adaptation locale du parasite, n'a pas encore été réalisé. De même, les bases génomiques de l'adaptation locale ont été

étudiées à l'aide de la génomique des populations et ont révélé des loci candidats associés à l'adaptation des parasites principalement à leur hôte. Cependant, les loci impliqués dans l'adaptation locale à l'environnement à la fois abiotique et biotique, au-delà des pathogènes fongiques des plantes, restent rares (Zhu et Zhao, 2007 ; Bonin, 2008 ; Stukenbrock et Bataillon, 2012 ; Simon et al., 2015 ; Dalongeville et al., 2018).

Les insectes parasites phytophages sont des modèles permettant d'étudier le processus de divergence. L'identification du nombre et de la nature des gènes (comme les gènes chimiosensoriels) et du rôle des régions répétées du génome (comme les éléments transposables, ou ET) dans la divergence adaptative des insectes phytophages (Gilbert et al., 2010 ; Simon et al., 2015 ; Shih et al., 2023), restent un sujet de recherche très étudié. Ainsi, des questions subsistent sur les liens entre les forces de sélection locales, la différenciation phénotypique adaptative des populations et la variation génomique adaptative sous-jacente à l'adaptation locale du parasite (Gloss et al., 2016). Il est donc nécessaire d'étudier les bases écologiques et génomiques de la divergence des insectes phytophages (Nooten et Hughes, 2017).

Parmi les insectes parasites phytophages, les pucerons constituent un modèle idéal pour comprendre les processus évolutifs impliqués dans la divergence des populations (Simon et al., 2015). Ces insectes présentent une parthénogenèse cyclique au cours de leur cycle de vie annuel. Ils ont un lien intime avec leur hôte (représentant le milieu biotique) mais aussi avec leur environnement (milieu abiotique). Par exemple, la transition entre les différentes étapes de leur cycle de vie dépend de la température, de l'humidité et de la photopériode. Les pucerons présentent une plasticité pour plusieurs caractères mais une spécificité d'hôte (Shih et al., 2023). Cette dépendance aux conditions abiotiques et biotiques, ainsi que l'identification d'interactions gène pour gène

lors de l'infestation (Züst et Agrawal, 2016), font des pucerons un modèle pertinent pour évaluer la contribution des conditions abiotiques et biotiques à l'adaptation hôte-parasite locale. De plus, les pucerons ont de petits génomes et se prêtent parfaitement aux manipulations expérimentales, ce qui en fait des modèles pour des tests expérimentaux de l'adaptation locale chez les parasites. Enfin, les pucerons font face à différentes pressions pour exploiter leur hôte. Par exemple, ils reconnaissent le message chimique produit par leur hôte spécifique (à l'aide de récepteurs chimiosensoriels). Ils rencontrent des barrières physiques et chimiques supplémentaires qui les empêchent de se nourrir, et l'hôte exerce des pressions écologiques à travers sa variation spatiale. Ces différentes pressions ont modelé des patrons d'adaptation locale à l'hôte chez les pucerons (aussi appelées races de pucerons).

Dysaphis plantaginea Passerini, ou puceron cendré du pommier, est un ravageur majeur des pommiers cultivés, provoquant d'importantes pertes économiques annuelles allant jusqu'à 30 %, provoquant des galles jaunâtres froissées sur les feuilles et des fruits petits et mal formés (Ismail et al., 2022). La répartition de *D. plantaginea* s'étend dans les régions tempérées (Amérique, Asie, Europe et Afrique du Nord ; Blackman et Eastop, 2000). Il existe des informations limitées sur l'écologie et l'évolution de *D. plantaginea* (Guillemaud et al., 2011 ; Miñarro et al., 2005 ; Pagliarani et al., 2016 ; Qubbaj et al., 2005 ; Wyss et al., 1999). La colonisation de *D. plantaginea* a probablement été rapide et récente. Elle suit la domestication du pommier cultivé (*Malus domestica* Borkh) importé par les Grecs en Europe depuis l'Asie Centrale, il y a environ 1500 ans (Cornille et al., 2014). La photopériode est un paramètre abiotique crucial pour l'éclosion, migration d'un hôte à un autre et l'induction de formes ailées et sexuées de *D. plantaginea*. Nous avons observé une variation des éclosions à travers l'Europe, avec des éclosions plus tardives au nord et plus précoces au sud. Cependant, il n'est pas clair si cette variation phénotypique des populations de *D. plantaginea*

reflète une adaptation locale à l'environnement abiotique ou biotique à la suite de sa récente colonisation de l'Europe. De plus, les génotypes de pommiers cultivés présentent différents degrés de tolérance à l'infestation de *D. plantaginea* (Miñarro et al., 2005 ; Pagliarani et al., 2016), et *D. plantaginea* infeste rarement le pommier sauvage européen, *Malus sylvestris* L. (Mill.). Le puceron cendré du pommier est donc un modèle idéal pour mieux comprendre les gènes, et les facteurs écologiques, impliqués dans la variation phénologique et potentiellement de l'adaptation locale du puceron cendré du pommier à son hôte (le pommier) ou à son environnement local.

Sous la direction de chercheurs experts dans le domaine et en utilisant des approches expérimentales, la génomique des populations et la génomique comparative, j'ai étudié les bases génomiques et écologiques impliqués dans la divergence adaptative de *D. plantaginea*. Premièrement, j'ai étudié l'évolution moléculaire des gènes chimiorécepteurs chez les pucerons (Chapitre 1), en lien avec l'adaptation à l'hôte. Deuxièmement, j'ai reconstruit l'histoire de la colonisation de *D. plantaginea* sur son hôte obligatoire, le pommier cultivé (Chapitre 2). Troisièmement, j'ai étudié les gènes impliqués dans l'adaptation locale de *D. plantaginea* à son environnement (Chapitre 3). Enfin, j'ai testé les bases écologiques de l'adaptation locale de *D. plantaginea* (Chapitre 4). Un résumé de chacun de ces chapitres est présenté ci-dessous.

Chapitre 1 : Dynamique évolutive des gènes chimiosensoriels et des éléments transposables chez les pucerons

L'identification des gènes impliqués dans l'adaptation des parasites (comme les insectes phytophages) à leur plante hôte fait l'objet de nombreuses recherches par les biologistes de l'évolution. Il existe également un intérêt croissant pour la compréhension de l'architecture génomique de l'adaptation des

insectes phytophages à leur hôte (Fouché et al., 2022 ; Oggenfuss et Croll, 2023), par exemple, avec l'étude du rôle des régions répétitives du génome, en particulier des éléments transposables (ET ; Gilbert et coll., 2021). Cependant, il existe peu d'informations sur ces gènes clés et la dynamique des ET impliquées dans l'adaptation des insectes phytophages à leur plante hôte.

L'olfaction et le goût ont un rôle majeur dans les interactions entre les insectes phytophages et leurs plantes hôtes. En effet, les insectes phytophages perçoivent les produits chimiques des plantes en activant des protéines spécifiques produites par des familles multigéniques chimiosensorielles (Simon et al., 2015 ; Robertson, 2019). Ces gènes chimiosensoriels sont très diversifiés (Santamaria et al., 2018) et jouent un rôle crucial dans l'adaptation et dans la spécialisation des insectes phytophages (Kambere et Lane, 2007 ; De Bruyne et Baker, 2008 ; Brand et al., 2015 ; Rinker et al., 2019). Parmi les gènes chimiosensoriels les plus importants chez les insectes, figurent les récepteurs olfactifs (OR) et les récepteurs gustatifs (GR) (Sanchez-Gracia et al., 2009 ; Robertson et al., 2019). Ces gènes sont connus pour évoluer par duplications de gènes et pseudogénéisation et de délétion ; Nimura et Nei et al., 2008 ; Sanchez-Garcia et al., 2009 ; Ramdya et Benton, 2010 ; Anderson et al., 2015 ; Benton, 2015). Les OR et GR sont connus pour avoir des taux d'évolution élevés par des événements d'expansion ou de contraction, qui dépendent des gènes chimiosensoriels et des espèces d'insectes (Halem et al., 2006 ; McBride, 2007 ; Joseph et Carlson, 2016 ; Sanchez-Gracia et al., 2009). Des études ont montré que l'évolution des gènes chimiosensoriels est régie par différentes pressions sélectives (sélection positive et purificatrice). Par exemple, des signatures de sélection ont été signalées dans les séquences des OR et GR du puceron du pois, suggérant des races d'hôtes en lien avec la spécialisation à l'hôte (Smadja et al., 2009 ; Meslin et al., 2022). De plus, les ET (Whiteman et Pierce, 2008 ; Meslin et al., 2022) font l'objet d'un intérêt grandissant. Des études sur le contenu en ET chez

de nombreuses espèces d'insectes ont révélé de fortes densités d'ET liées aux événements de duplication (Gilbert et al., 2021). Cependant, les processus évolutifs impliqués dans l'évolution des gènes OR et GR et le rôle des ET dans l'adaptation des insectes phytophages à leur hôte restent floue.

Les pucerons présentent différentes caractéristiques permettant d'étudier l'évolution des gènes clés impliqués dans l'adaptation à l'hôte. Ils ont une grande diversité génétique et phénotypique intraspécifique (Loxdale, 2008 ; Zytynska et al., 2013 ; Leclair et al., 2016). Ils ont également une gamme de préférences en matière d'hôtes, des cycles de vie complexes présentant un polyphénisme et différents cycles de reproduction sexuée et asexuée (Srinivasan et Brisson., 2012). Les pucerons peuvent développer leur cycle de vie sur un hôte particulier ou sur plusieurs espèces hôtes appartenant à une famille de plante unique ou non apparentée (Dixon, 1977 ; Williams et Dixon., 2007). Leur système chimiosensoriel est essentiel pour détecter les produits chimiques et leur permettre d'obtenir des informations sur l'environnement abiotique et biotique (c'est-à-dire l'hôte ; Robertson et al., 2019). Ce système est crucial pour l'adaptation des pucerons aux différents environnements et hôtes. Les études sur les génomes de pucerons disponibles ont révélé des duplications de gènes en lien avec la diversification des pucerons (The International Aphid Genomics Consortium, 2010 ; Mathers et al., 2017 ; Thorpe et al., 2018 ; Fernandez et al., 2020 ; Julca et al., 2020 ; Biello et coll., 2021). Cette dynamique du génome, ainsi que l'évolution rapide des gènes et l'enrichissement de gènes dotés de fonctions clés en lien avec l'adaptation de l'hôte (par exemple, les gènes des récepteurs chimio sensoriels), sont à la base de la spécialisation des pucerons vers leur hôte (Smadja et al., 2009 ; Puinean et al., 2010 ; Nicholson et al., 2015 ; Quan et al., 2019).

Dans ce chapitre, nous avons étudié l'évolution des gènes chimio sensoriels au cours de l'histoire évolutive des pucerons, et le rôle des ET dans leur évolution. Pour réaliser cette étude, 11 espèces de pucerons, et une espèce apparentée, présentant différentes préférences écologiques, ont été confrontées à l'aide d'outils de génomique comparative. Nous avons inclus et présenté la version 3 du génome (assemblage génomique de haute qualité) de *Dysaphis plantaginea*, assemblé dans l'équipe. Nous avons annoté manuellement 521 OR et 399 GR parmi neuf espèces de pucerons à l'aide de trois logiciels différents : Exonerate (Slater et Birney, 2005), le pipeline InsectOR (Karpe et al., 2021) et Scipio (Keller et al., 2008). Nous avons inclus 147 OR et 181 GR déjà publiés d'*Acyrtosiphon pisum* Harris, *Aphis glycines* Matsumura et *Daktulosphaira vitifoliae* Fitch (Robertson et al., 2019 ; Rispe et al., 2020), aucun gène supplémentaire n'ayant été identifié. Nous avons constaté que la principale pression sélective à l'origine de l'évolution des gènes chimio sensoriels des pucerons était la sélection purificatrice. Nous avons aussi montré que l'évolution des gènes OR et GR suit le modèle de naissance et de mort, et évoluant par répétitions en tandem. Nous montrons que les ET ne sont pas impliqués dans la diversité et l'évolution des OR and GR. Cependant, nous avons observé des dynamiques évolutive espèce-spécifiques, avec les espèces de pucerons infestant les fruitiers montrant une activité très récente des ET.

Cette étude a permis d'obtenir des informations sur l'évolution moléculaire des gènes fondamentaux impliqués dans l'adaptation des pucerons à leurs hôtes. Ce chapitre permet de mieux comprendre les processus évolutifs en jeu lors de l'évolution des pucerons à large échelle évolutive. Ce travail est en cours de soumission à la revue *Molecular Biology and Evolution*.

Chapitre 2 : Historique de la colonisation du puceron cendré du pommier

La compréhension des processus évolutifs à l'origine de la colonisation des espèces et de l'expansion de leur aire de répartition est une question fondamentale dans le domaine de la biologie évolutive (Austerlitz et al., 1997 ; Excoffier et al., 2009 ; Rius et Darling, 2014 ; Hoffman et Courchamp, 2016 ; Angert et coll., 2020). Les parasites des cultures constituent un modèle parfait pour étudier le processus sous-jacent à une colonisation rapide et à l'expansion de l'aire de répartition, car ces parasites présentent de fréquents changements d'hôtes (Stuckenbrock et McDonald, 2008 ; Gladieux et al., 2014 ; Garnas et al., 2016). L'évaluation de la diversité des parasites des cultures et de la structure des populations est essentielle pour comprendre la colonisation des espèces ainsi que l'expansion de l'aire de répartition et pour étudier plus en détail les gènes clés impliqués dans l'adaptation (Tiffin et Ross-Ibarra, 2014).

Nous avons donc voulu comprendre l'histoire de la colonisation de *Dysaphis plantaginea*, l'un des principaux ravageurs du pommier cultivé en Europe, en Amérique du Nord et au Moyen-Orient, encore peu documentée (Guillemaud et al., 2011). Pourrions-nous trouver un patron typique (contraction suivie d'une expansion des populations) associé à la colonisation d'un nouvel environnement chez *D. plantaginea*? Nous voulions également déterminer s'il y avait un flux de gènes entre les populations lors de la colonisation par *D. plantaginea*, car les pucerons présentent de très grandes capacités de dispersion (Roderick, 1996). Nous nous sommes également demandé si d'autres acteurs étaient importants dans l'histoire de la colonisation de *D. plantaginea*, par exemple les bactéries endosymbiotiques hébergées par les pucerons. Ces endosymbiontes sont connus pour jouer un rôle dans la colonisation de nouveaux habitats (Tsuchida et al., 2002 ; Russell et al., 2013 ; Zepeda-Paulo et al., 2018 ; Lenhart et White, 2020 ; Leclair et al., 2021).

Ce chapitre présente l'utilisation de marqueurs moléculaires pour étudier l'histoire démographique de *D. plantaginea* en Afrique du Nord, en Amérique du Nord et en Europe. Nous avons utilisé des marqueurs mitochondriaux (*CytB* et *CO1*), bactériens (*ARNr 16s* et *TrnpB*) et 30 marqueurs microsatellites nouvellement développés. Nous avons utilisé plus de 650 colonies de pucerons cendré du pommier génotypées dans 52 vergers répartis dans 13 pays. Nous avons évalué sa diversité génétique, déterminé la structure des populations et caractérisé la diversité de la communauté bactérienne endosymbiotique. Nous avons détecté cinq populations de pucerons cendré du pommier isolées par la distance, faiblement différenciées génétiquement, et échangeant de forts flux de gènes. Nos résultats suggèrent aussi une colonisation de *D. plantaginea* de l'Europe de l'Est vers l'Europe de l'Ouest, suivie d'une colonisation de l'Amérique du Nord et du Maroc. Nous avons également tenté de trouver la population source de colonisation en utilisant les approches ABC-RF. Cependant, le flux de gènes substantiels entre populations et le manque d'échantillons en provenance du Moyen-Orient ont entravé ces analyses. Nous n'avons trouvé aucun rôle central d'endosymbiontes secondaires lors de la colonisation du puceron cendré du pommier, et nous avons confirmé la présence de la bactérie endosymbiotique obligatoire *Buchnera aphidicola* Munson dans toutes les populations de pucerons de la zone d'échantillonnage.

Cette étude a permis de démontrer une expansion récente de l'aire de répartition d'un puceron ravageur des arbres fruitiers, impliquant de forts flux de gènes entre populations très diversifiées génétiquement. Les conséquences de tels flux de gènes entre les populations du puceron cendré du pommier restent à explorer avec des marqueurs génomiques supplémentaires (par exemple, le polymorphisme nucléotidique unique ou « SNP »). Les travaux du chapitre 2, reconstruisant l'histoire évolutive du puceron cendré du pommier à l'aide de marqueurs microsatellites, ont été publiés dans la *Peers Community in Evolutionary*

Biology (<https://doi.org/10.24072/pcjournal.26>), et ces travaux ont également été vulgarisés dans un article dans la revue en ligne « Passion Entomologie » (<https://passion-entomologie.fr/pommier-puceron-cendre/>).

Chapitre 3 : Bases génomiques de l'adaptation locale du puceron cendré du pommier à son hôte

Les régions génomiques très divergentes entre populations issues de différents environnements peuvent résulter d'adaptation locale (Andolfatto, 2001 ; Nielsen, 2005). La propagation d'allèles avantageux dans une population par d'un balayage sélectif peut expliquer ces régions génomiques divergentes. (Stephan, 2019). Les agents pathogènes et les ravageurs des cultures sont des modèles pour comprendre les gènes impliqués dans l'adaptation locale (Stukenbrock et Bataillon, 2012 ; Badouin et al., 2017 ; Chen et al., 2021 ; Cao et al., 2022). La plupart des études ont été réalisées sur des espèces modèles de ravageurs des cultures. Chez *Acyrtosiphon pisum*, des gènes candidats sous sélection positive codant pour des protéines de type *cathepsine B* ont été détectés. Ces gènes ont une activité protéolytique pour faire face à la réponse de la plante (Review in Simon et al., 2015). De la même manière, des gènes chimio sensoriels montrent des signatures de sélection positive liée à la spécialisation de l'hôte (Smadja et al., 2009). Les analyses du génome de différents biotypes de pucerons du soja (*Aphis glycines*) ont permis de détecter un balayage sélectif dans les gènes codant pour des effecteurs (Coates et al., 2020). Cependant, il n'existe aucune étude sur les bases génomiques d'adaptation locale du puceron cendré du pommier. De plus, les informations sur les régions génomiques et les pressions sélectives sous-jacentes à l'adaptation locale ne sont pas clarifiées chez les insectes ravageurs.

Dans ce chapitre 3, nous avons étudié les bases génomiques de l'adaptation (locale) du puceron cendré du pommier. Nous avons utilisé notre première observation sur l'histoire de la colonisation du puceron cendré du pommier (Olvera-Vazquez et al., 2021 (chapitre 2)) pour sélectionner des individus représentatifs de la structure des populations et de la diversité génétique, et nous avons ajouté des individus du Canada et d'Israël. Nous avons également inclus des échantillons des espèces sœurs *Dysaphis devectora* Walker (Roumanie et Iran) et *Dysaphis pyri* Boyer de Fonscolombe (France, Espagne et Iran). Nous avons extrait l'ADN de chaque puceron et les avons envoyés pour un reséquençage des extrémités appariées du génome entier (2*150 pb, 30X) à l'aide de la plateforme Illumina (NovaSeq6000). A l'aide de SNPs, nous avons essayé de répondre aux questions suivantes : Est-ce que les trois espèces de pucerons forment des groupes génétiques distincts? Quelle était la diversité génétique et la structure de la population chez *D. plantaginea* ? Répondre ces questions est fondamental pour éviter la détecter de faux positifs dans les gènes candidats impliqués dans l'adaptation (Li et al., 2012 ; Tiffin et Ross-Ibarra, 2014 ; Hoban et al., 2016). Enfin, observe-t-on des signatures de sélection positive dans les génomes de chaque population qui pourrait suggérer une adaptation ou une adaptation locale chez *Dysaphis* spp.?

Les trois espèces de pucerons, *D. devectora*, *D. plantaginea* et *D. pyri*, forment trois groupes génétiques distincts. Chez *D. plantaginea*, nous avons trouvé quatre groupes génétiques : nord-américain (Canada et États-Unis), israélien, marocain et l'Espagne et un autre comprenant le reste des pays européens. Nous avons également constaté une expansion récente des populations des groupes génétiques de *D. plantaginea*. En plus du flux génétique entre le groupe génétique Amérique et Maroc/Espagne et Europe et Maroc/Espagne. Nous avons ensuite

scanné les génomes des six populations pour des patrons de sélection positive chez *D. devectora*, *D. pyri*, *D. plantaginea* d'Amérique d'Israël, du Maroc et des pays européens. Nous avons trouvé des signaux de balayages sélectifs en utilisant deux statistiques (Alachiotis et Pavlidis, 2018) (Alachiotis et al., 2012). Nous avons complété ces analyses par une analyses des bases écologiques de l'adaptation locale chez les pucerons (Chapitre 4).

Chapitre 4 : Test expérimental de l'adaptation locale du puceron cendré du pommier (*Dysaphis plantaginea*) lors de sa récente colonisation rapide sur son hôte, le pommier cultivé *Malus domestica* en Europe

Les interactions plantes-parasites sont utilisées comme modèle pour étudier les processus évolutifs impliqués dans l'adaptation locale (Garrido et al., 2011). Les variations de sélection entre les environnements peuvent conduire à l'adaptation locale du parasite à son hôte (Kawecki et Ebert, 2004 ; Olazcuaga et al., 2022). Des jardins communs ont été utilisés pour identifier les caractéristiques clés qui sous-tendent l'adaptation locale et les facteurs écologiques qui déterminent l'adaptation (Kaltz et Shykoff, 1998 ; Tiffin et Ross-Ibarra, 2014 ; Greischar et Koskella, 2007 ; Nuismar et Gandon, 2008 ; Nooten et Hughes, 2017). Des patrons d'adaptation locale ont été découverts à l'aide d'expériences d'infection croisée (Kaltz et Shykoff, 1998 ; Kawecki et Ebert, 2004 ; Greischar et Koskella, 2007 ; Sicard et al., 2007), et dans d'autre cas, des patrons de maladaptation à leurs hôtes (Kaltz et al., 1999). Cependant, les contributions de l'environnement abiotique (c'est-à-dire l'environnement local) et biotique (c'est-à-dire la plante hôte locale) à l'adaptation locale du parasite sont encore inconnues. L'étude des contributions de l'environnement abiotique et biotique local permet de comprendre comment un nouvel environnement peut favoriser l'émergence de parasites conduisant à l'extinction de l'hôte, prédire les épidémies et

comprendre la migration des vecteurs vers de nouvelles niches écologiques. Pour ces raisons, il est fondamental de réaliser des expériences de transplantation et d'infection croisée multi-populations pour tester l'effet des environnements biotiques et abiotiques reflétant un cline spatial (Gandon et al., 2008 ; Brown et Tellier, 2011). Les tests expérimentaux disponibles d'adaptation locale aux plantes hôtes se sont concentrés principalement sur les phytopathogènes (champignons principalement). En effet, les champignons pathogènes sont plus faciles à manipuler pour contrôler les infections que d'autres grands groupes de parasites, comme les insectes phytophages. Les expériences d'infection croisée et de transplantation représentent un défi technique et en ressources humaines nécessaires ; il est donc moins fréquent de trouver des études à ce sujet. L'une des motivations de ce chapitre est d'utiliser des approches expérimentales pour mieux comprendre les processus évolutifs derrière l'adaptation locale des insectes phytophages aux conditions biotiques et abiotiques, ce qui est essentiel au développement de stratégies de lutte antiparasitaire (Kawecki et Ebert, 2004 ; Tiffin et Ross-Ibarra, 2014).

Au cours de ce chapitre, nous avons étudié l'étendue de l'adaptation locale des insectes phytophages à leur environnement abiotique local et biotique (le pommier) en utilisant les populations de *D. plantaginea* comme modèles. Nous avons exploré les réponses de différents génotypes de pommiers cultivés à l'infestation de génotypes locaux et étrangers de pucerons cendré du pommier d'origines différentes. L'expérience consistait en trois jardins communs représentant un gradient environnemental : Espagne, France et Belgique. Nous avons réalisé une expérience d'infestation croisée de pucerons du début du printemps au début de l'été 2021. Nous avons utilisé 28 génotypes de pommiers comprenant 15 génotypes de pommiers cultivés (cinq génotypes locaux de chaque pays), neuf génotypes de pommiers sauvages (*Malus sylvestris*) et quatre génotypes de pommiers avec différents niveaux de sensibilité face aux

infestations de pucerons : trois génotypes de pommiers tolérants (un 'Priscilla', un 'Florina' et un génotype de l'espèce ornementale *Malus floribunda* Siebold ex Van Houtte), et un génotype sensible, *Malus domestica* Golden Delicious. Nous avons échantillonné neuf génotypes représentant trois populations de pucerons cendrés du pommier d'un pays et les avons échangés entre les trois jardins communs au printemps 2021. Nous avons placé trois pucerons adultes sur neuf points d'infestation différents (représentés par une feuille sélectionnée au hasard) sur chaque génotype de pommier et les avons protégés avec une cage à pucerons développée dans l'équipe ECLECTIC. Après 5 à 7 jours, nous avons récupéré les colonies de pucerons et pris une photo de chaque colonie au laboratoire. Le plan expérimental du chapitre 4 a été publié dans Peer Community in Ecology pour améliorer notre expérience sur la base des suggestions de chercheurs possédant une grande expertise dans le domaine (https://forgemia.inra.fr/amandine.cornille/local_adaptation_dp).

Nous avons déterminé la valeur adaptative des colonies de pucerons en obtenant le nombre d'individus dans chaque colonie, le nombre de stades différents et la surface de chaque puceron. Nous avons créé une classification utilisant la morphologie des pucerons de chaque stade de développement basé sur l'intelligence artificielle. Nous avons donc utilisé des outils d'apprentissage automatique pour classer chaque stade de puceron à partir de l'image de la colonie échantillonnée sur le terrain en collaboration avec le laboratoire de physique de l'université d'Angers, dirigé par le Dr David Rousseau, et développé avec l'ingénieur de recherche Herearii Metuarea.

Nous avons dû réduire notre univers de données, en éliminant le site belge et les génotypes en raison du faible nombre de points d'infestation représentatifs dans ce jardin commun dû à la complexité de l'expérimentation. Ensuite, nous nous sommes concentrés uniquement sur les sites français et espagnols. Nos

résultats préliminaires ont montré une interaction entre la localisation commune du jardin et l'origine des pucerons (France et Espagne) sur le fitness des pucerons. Nous avons en outre étudié les différences de taux de croissance des colonies dans chaque jardin commun et avons montré que cette dernière interaction s'expliquait principalement par les différences de taux de croissance dans le site français entre les populations de pucerons espagnoles et françaises. Nous avons également comparé la condition physique de chaque population de pucerons parmi les jardins communs et avons montré des différences de condition physique de la population française de pucerons, montrant une meilleure condition physique dans son jardin commun étranger. Ces résultats préliminaires suggèrent une maladaptation du puceron français.

References

- Ahmad, A., Wallau, G. L., & Ren, Z. (2021). Characterization of Mariner transposons in seven species of Rhus gall aphids. *Scientific reports*, 11(1), 1-11. <https://doi.org/10.1038/s41598-021-95843-5>
- Alachiotis, N., Stamatakis, A., & Pavlidis, P. (2012). OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics*, 28(17), 2274-2275. <https://doi.org/10.1093/bioinformatics/bts419>
- Alachiotis, N., & Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications biology*, 1(1), 79. <https://doi.org/10.1038/s42003-018-0085-8>
- Andersson, M. N., Löfstedt, C., & Newcomb, R. D. (2015). Insect olfaction and the evolution of receptor tuning. *Frontiers in Ecology and Evolution*, 3, 53. <https://doi.org/10.3389/fevo.2015.00053>
- Andolfatto, P. (2001). Adaptive hitchhiking effects on genome variability. *Current opinion in genetics & development*, 11(6), 635-641. [https://doi.org/10.1016/S0959-437X\(00\)00246-X](https://doi.org/10.1016/S0959-437X(00)00246-X)
- Angert, A. L., Bontrager, M. G., & Ågren, J. (2020). What do we really know about adaptation at range edges?. *Annual Review of Ecology, Evolution, and Systematics*, 51, 341-361. <https://doi.org/10.1146/annurev-ecolsys-012120-091002>
- Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P. H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical population biology*, 51(2), 148-164. <https://doi.org/10.1006/tpbi.1997.1302>
- Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguilera, G., Snirc, A., ... & Giraud, T. (2017). Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Molecular Ecology*, 26(7), 2041-2062. <https://doi.org/10.1111/mec.13976>
- Benton, R., Sachse, S., Michnick, S. W., & Vosshall, L. B. (2006). Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. *PLoS biology*, 4(2), e20. <https://doi.org/10.1371/journal.pbio.0040020>

Benton, R. (2015). Multigene family evolution: perspectives from insect chemoreceptors. *Trends in ecology & evolution*, 30(10), 590-600. <https://doi.org/10.1016/j.tree.2015.07.009>

Biello, R., Singh, A., Godfrey, C. J., Fernández, F. F., Mugford, S. T., Powell, G., ... & Mathers, T. C. (2021). A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Molecular ecology resources*, 21(1), 316-326. <https://doi.org/10.1111/1755-0998.13258>

Blackman, R. L., & Eastop, V. F. (2000). *Aphids on the world's crops: an identification and information guide* (No. Ed. 2). John Wiley & Sons Ltd

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology letters*, 16(9), 1195-1205. <https://doi.org/10.1111/ele.12150>

Brand, P., Ramírez, S. R., Leese, F., Quezada-Euan, J. J. G., Tollrian, R., & Eltz, T. (2015). Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). *BMC evolutionary biology*, 15(1), 1-16. <https://doi.org/10.1186/s12862-015-0451-9>

Bonin, A. (2008). Population genomics: a new generation of genome scans to bridge the gap with functional genomics. <https://doi.org/10.1111/j.1365-294X.2008.03854.x>

Bouallègue, M., Filée, J., Kharrat, I., Mezghani-Khemakhem, M., Rouault, J. D., Makni, M., & Capy, P. (2017). Diversity and evolution of mariner-like elements in aphid genomes. *BMC genomics*, 18(1), 1-12. <https://doi.org/10.1186/s12864-017-3856-6>

Brown, J. K., & Tellier, A. (2011). Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual review of phytopathology*, 49, 345-367. <https://doi.org/10.1146/annurev-phyto-072910-095301>

Cao, L. J., Song, W., Chen, J. C., Fan, X. L., Hoffmann, A. A., & Wei, S. J. (2022). Population genomic signatures of the oriental fruit moth related to the Pleistocene climates. *Communications Biology*, 5(1), 142.

<https://doi.org/10.1038/s42003-022-03097-2>

Capblancq, T., Lachmuth, S., Fitzpatrick, M. C., & Keller, S. R. (2023). From common gardens to candidate genes: exploring local adaptation to climate in red spruce. *New Phytologist*, 237(5), 1590-1605. <https://doi.org/10.1111/nph.18465>

Cornille, A., Giraud, T., Smulders, M. J., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, 30(2), 57-65. <https://doi.org/10.1016/j.tig.2013.10.002>

Chen, Y., Liu, Z., Régnière, J., Vasseur, L., Lin, J., Huang, S., ... & You, S. (2021). Large-scale genome-wide study reveals climate adaptive variability in a cosmopolitan pest. *Nature Communications*, 12(1), 7206. <https://doi.org/10.1038/s41467-021-27510-2>

Coates, B. S., Hohenstein, J. D., Giordano, R., Donthu, R. K., Michel, A. P., Hodgson, E. W., & O'Neal, M. E. (2020). Genome scan detection of selective sweeps among biotypes of the soybean aphid, *Aphis glycines*, with differing virulence to resistance to *A. glycines* (Rag) traits in soybean, *Glycine max*. *Insect Biochemistry and Molecular Biology*, 124, 103364. <https://doi.org/10.1016/j.ibmb.2020.103364>

Dalongeville, A., Benestan, L., Mouillot, D., Lobreaux, S., & Manel, S. (2018). Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). *BMC genomics*, 19, 1-13. <https://doi.org/10.1186/s12864-018-4579-z>

De Bruyne, M., & Baker, T. C. (2008). Odor detection in insects: volatile codes. *Journal of chemical ecology*, 34(7), 882-897. <https://doi.org/10.1007/s10886-008-9485-4>

Dixon, A. F. G. (1977). Aphid ecology: life cycles, polymorphism, and population regulation. *Annual Review of Ecology and Systematics*, 329-353. <https://www.jstor.org/stable/2096732>

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481-501. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173414>

Fernández, R., Marcet-Houben, M., Legeai, F., Richard, G., Robin, S., Wucher, V., ... & Tagu, D. (2020). Selection following gene duplication shapes recent genome evolution in the pea aphid *Acyrtosiphon pisum*. *Molecular biology and evolution*, 37(9), 2601-2615. <https://doi.org/10.1093/molbev/msaa110>

Fouché, S., Oggenfuss, U., Chanclud, E., & Croll, D. (2022). A devil's bargain with transposable elements in plant pathogens. *Trends in Genetics*, 38(3), 222-230. <https://doi.org/10.1016/j.tig.2021.08.005>

Gandon, S., Buckling, A., Decaestecker, E., & Day, T. (2008). Host-parasite coevolution and patterns of adaptation across time and space. *Journal of evolutionary biology*, 21(6), 1861-1866. <https://doi.org/10.1111/j.1420-9101.2008.01598.x>

Garnas, R., J., Auger-Rozenberg, M. A., Roques, A., Bertelsmeier, C., Wingfield, M. J., Saccaggi, D. L., ... & Slippers, B. (2016). Complex patterns of global spread in invasive insects: eco-evolutionary and management consequences. *Biological Invasions*, 18, 935-952. <https://doi.org/10.1007/s10530-016-1082-9>

Garrido, E., Andraca-Gómez, G., & Fornoni, J. (2012). Local adaptation: simultaneously considering herbivores and their host plants. *New Phytologist*, 193(2), 445-453. <https://doi.org/10.1111/j.1469-8137.2011.03923.x>

Gianoli, E., & Valladares, F. (2012). Studying phenotypic plasticity: the advantages of a broad approach. *Biological Journal of the Linnean Society*, 105(1), 1-7. <https://doi.org/10.1111/j.1095-8312.2011.01793.x>

Gilbert, C., Peccoud, J., & Cordaux, R. (2021). Transposable elements and the evolution of insects. *Annual Review of Entomology*, 66(1), 355-372. <https://doi.org/10.1146/annurev-ento-070720-074650>

Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguilera, G., De Vienne, D. M., ... & Giraud, T. (2014). Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular ecology*, 23(4), 753-773. <https://doi.org/10.1111/mec.12631>

Gladieux, P., Wilson, B. A., Perraudeau, F., Montoya, L. A., Kowbel, D., Hann-Soden, C., ... & Taylor, J. W. (2015). Genomic sequencing reveals historical, demographic

and selective factors associated with the diversification of the fire-associated fungus *Neurospora discreta*. *Molecular ecology*, 24(22), 5657-5675. <https://doi.org/10.1111/mec.13417>

Gloss, A. D., Groen, S. C., & Whiteman, N. K. (2016). A genomic perspective on the generation and maintenance of genetic diversity in herbivorous insects. *Annual review of ecology, evolution, and systematics*, 47, 165-187. <https://doi.org/10.1146/annurev-ecolsys-121415-032220>

Greischar, M. A., & Koskella, B. (2007). A synthesis of experimental work on parasite local adaptation. *Ecology letters*, 10(5), 418-434. <https://doi.org/10.1111/j.1461-0248.2007.01028.x>

Guillemaud, T., Blin, A., Simon, S., Morel, K., & Franck, P. (2011). Weak spatial and temporal population genetic structure in the rosy apple aphid, *Dysaphis plantaginea*, in French apple orchards. *Plos one*, 6(6), e21263. <https://doi.org/10.1371/journal.pone.0021263>

Hallem, E. A., Dahanukar, A., & Carlson, J. R. (2006). Insect odor and taste receptors. *Annu. Rev. Entomol.*, 51, 113-135. <https://doi.org/10.1146/annurev.ento.51.051705.113646>

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., ... & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4), 379-397. <https://doi.org/10.1086/688018>

Hoffmann, B. D., & Courchamp, F. (2016). Biological invasions and natural colonisations: are they that different? *NeoBiota*, 29, 1-14. <https://doi.org/10.3897/neobiota.29.6959>

International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS biology*, 8(2), e1000313. <https://doi.org/10.1371/journal.pbio.1000313>

Joseph, R. M., & Carlson, J. R. (2015). *Drosophila* chemoreceptors: a molecular interface between the chemical world and the brain. *Trends in Genetics*, 31(12), 683-695. <https://doi.org/10.1016/j.tig.2015.09.005>

Julca, I., Marcet-Houben, M., Cruz, F., Vargas-Chavez, C., Johnston, J. S., Gómez-Garrido, J., ... & Gabaldón, T. (2020). Phylogenomics identifies an ancestral burst of gene duplications predating the diversification of Aphidomorpha. *Molecular Biology and Evolution*, 37(3), 730-756. <https://doi.org/10.1093/molbev/msz261>

Kaltz, O., & Shykoff, J. A. (1998). Local adaptation in host-parasite systems. *Heredity*, 81(4), 361-370. <https://doi.org/10.1046/j.1365-2540.1998.00435.x>

Kaltz, O., Gandon, S., Michalakis, Y., & Shykoff, J. A. (1999). Local maladaptation in the anther-smut fungus *Microbotryum violaceum* to its host plant *Silene latifolia*: evidence from a cross-inoculation experiment. *Evolution*, 53(2), 395-407. <https://doi.org/10.1111/j.1558-5646.1999.tb03775.x>

Kambere, M. B., & Lane, R. P. (2007). Co-regulation of a large and rapidly evolving repertoire of odorant receptor genes. *BMC neuroscience*, 8(3), 1-16. <https://doi.org/10.1186/1471-2202-8-S3-S2>

Karvonen, A., & Seehausen, O. (2012). The role of parasitism in adaptive radiations—when might parasites promote and when might they constrain ecological speciation?. *International Journal of Ecology*, 2012. <https://doi.org/10.1155/2012/280169>

Karpe, S. D., Tiwari, V., & Ramanathan, S. (2021). InsectOR—Webserver for sensitive identification of insect olfactory receptor genes from non-model genomes. *Plos one*, 16(1), e0245324. <https://doi.org/10.1371/journal.pone.0245324>

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology letters*, 7(12), 1225-1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>

Keller, O., Odronitz, F., Stanke, M., Kollmar, M., & Waack, S. (2008). Scipio: using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC bioinformatics*, 9(1), 1-12. <https://doi.org/10.1186/1471-2105-9-278>

Leclair, M., Pons, I., Mahéo, F., Morlière, S., Simon, J. C., & Outreman, Y. (2016).

Diversity in symbiont consortia in the pea aphid complex is associated with large phenotypic variation in the insect host. *Evolutionary Ecology*, 30(5), 925-941. <https://doi.org/10.1007/s10682-016-9856-1>

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PloS one*, 3(12), e4010. <https://doi.org/10.1371/journal.pone.0004010>

Lenhart PA, White JA (2020) Endosymbionts facilitate rapid evolution in a polyphagous herbivore. *Journal of Evolutionary Biology*, 33, 1507–1511. <https://doi.org/10.1111/jeb.13697>

Leonardo, T. E., & Mondor, E. B. (2006). Symbiont modifies host life-history traits that affect gene flow. *Proceedings of the Royal Society B: Biological Sciences*, 273(1590), 1079-1084. <https://doi.org/10.1098/rspb.2005.3408>

Lenormand, T. (2012). From local adaptation to speciation: specialization and reinforcement. *International Journal of Ecology*, 2012. <https://doi.org/10.1155/2012/508458>

Li, J., Li, H., Jakobsson, M., Li, S. E. N., SjöDin, P. E. R., & Lascoux, M. (2012). Joint analysis of demography and selection in population genetics: where do we stand and where could we go?. *Molecular ecology*, 21(1), 28-44. <https://doi.org/10.1111/j.1365-294X.2011.05308.x>

Loxdale, H. D. (2008). The nature and reality of the aphid clone: genetic variation, adaptation and evolution. *Agricultural and Forest Entomology*, 10(2), 81-90. <https://doi.org/10.1111/j.1461-9563.2008.00364.x>

Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., Baa-Puyoulet, P., ... & Hogenhout, S. A. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome biology*, 18(1), 1-20. <https://doi.org/10.1186/s13059-016-1145-3>

McBride, C. S. (2007). Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proceedings of the National Academy of Sciences*, 104(12), 4996-5001. <https://doi.org/10.1073/pnas.0608424104>

Meek, M. H., Beever, E. A., Barbosa, S., Fitzpatrick, S. W., Fletcher, N. K., Mittan-

Moreau, C. S., ... & Hellmann, J. J. (2023). Understanding local adaptation to prepare populations for climate change. *BioScience*, 73(1), 36-47. <https://doi.org/10.1093/biosci/biac101>

Meslin, Camille, Pauline Mainet, Nicolas Montagné, Stéphanie Robin, Fabrice Legeai, Anthony Bretaudeau, J Spencer Johnston, et al. 2022. "Spodoptera Littoralis Genome Mining Brings Insights on the Dynamic of Expansion of Gustatory Receptors in Polyphagous Noctuidae." *G3 Genes|Genomes|Genetics*. <https://doi.org/10.1093/g3journal/jkac131>

Miñarro, M., Hemptinne, J. L., & Dapena, E. (2005). Colonization of apple orchards by predators of *Dysaphis plantaginea*: sequential arrival, response to prey abundance and consequences for biological control. *BioControl*, 50, 403-414. <https://doi.org/10.1007/s10526-004-5527-1>

Nicholson, S. J., Nickerson, M. L., Dean, M., Song, Y., Hoyt, P. R., Rhee, H., ... & Puterka, G. J. (2015). The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC genomics*, 16(1), 1-16. <https://doi.org/10.1186/s12864-015-1525-1>

Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39, 197-218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>

Niimura, Y., & Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PloS one*, 2(8), e708. <https://doi.org/10.1371/journal.pone.0000708>

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in ecology & evolution*, 26(4), 160-167. <https://doi.org/10.1016/j.tree.2011.01.001>

Nooten, S. S., & Hughes, L. (2017). The power of the transplant: direct assessment of climate change impacts. *Climatic Change*, 144, 237-255. <https://doi.org/10.1007/s10584-017-2037-6>

Oggenfuss, U., & Croll, D. (2023). Recent transposable element bursts are associated with the proximity to genes in a fungal plant pathogen. *PLoS Pathogens*, 19(2), e1011130. <https://doi.org/10.1371/journal.ppat.1011130>

Olazcuaga, L., Foucaud, J., Deschamps, C., Loiseau, A., Claret, J. L., Vedovato, R., ... & Estoup, A. (2022). Rapid and transient evolution of local adaptation to seasonal host fruits in an invasive pest fly. *Evolution Letters*, 6(6), 490-505. <https://doi.org/10.1002/evl3.304>

Olvera-Vazquez, S. G., Remoué, C., Venon, A., Rousselet, A., Grandcolas, O., Azrine, M., ... & Cornille, A. (2021). Large-scale geography survey provides insights into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa. *Peer Community Journal*, 1. <https://doi.org/10.24072/pcjournal.26>

Pagliarani, G., Dapena, E., Miñarro, M., Denancé, C., Lespinasse, Y., Rat-Morris, E., ... & Tartarini, S. (2016). Fine mapping of the rosy apple aphid resistance locus Dp-fl on linkage group 8 of the apple cultivar 'Florina'. *Tree Genetics & Genomes*, 12, 1-12. <https://doi.org/10.1007/s11295-016-1015-x>

Peccoud, Jean, & Simon, Jean-Christophe. (2010). The pea aphid complex as a model of ecological speciation. *Ecological Entomology*, 35, 119-130. <https://doi.org/10.1111/j.1365-2311.2009.01147.x>

Powell, J. R., & Tabachnick, W. J. (2013). History of domestication and spread of *Aedes aegypti*-a review. *Memórias do Instituto Oswaldo Cruz*, 108, 11-17. <https://doi.org/10.1590/0074-0276130395>

Puinean, A. M., Foster, S. P., Oliphant, L., Denholm, I., Field, L. M., Millar, N. S., ... & Bass, C. (2010). Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLOS genetics*, 6(6), e1000999. <https://doi.org/10.1371/journal.pgen.1000999>

Quan, Q., Hu, X., Pan, B., Zeng, B., Wu, N., Fang, G., ... & Zhan, S. (2019). Draft genome of the cotton aphid *Aphis gossypii*. *Insect Biochemistry and Molecular Biology*, 105, 25-32. <https://doi.org/10.1016/j.ibmb.2018.12.007>

Qubbaj, T., Reineke, A., & Zebitz, C. P. W. (2005). Molecular interactions between rosy apple aphids, *Dysaphis plantaginea*, and resistant and susceptible cultivars of its primary host *Malus domestica*. *Entomologia Experimentalis et Applicata*, 115(1), 145-152. <https://doi.org/10.1111/j.1570-7458.2005.00255.x>

Ramdaya, P., & Benton, R. (2010). Evolving olfactory systems on the fly. *Trends in Genetics*, 26(7), 307-316. <https://doi.org/10.1016/j.tig.2010.04.004>

Razzaque, S., Heckman, R. W., & Juenger, T. E. (2023). Seed size variation impacts local adaptation and life-history strategies in a perennial grass. *Proceedings of the Royal Society B*, 290(1998), 20222460. <https://doi.org/10.1098/rspb.2022.2460>

Rinker, D. C., Specian, N. K., Zhao, S., & Gibbons, J. G. (2019). Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. *Proceedings of the National Academy of Sciences*, 116(27), 13446-13451. <https://doi.org/10.1073/pnas.1901093116>

Rispe, C., Legeai, F., Nabity, P. D., Fernández, R., Arora, A. K., Baa-Puyoulet, P., ... & Tagu, D. (2020). The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC biology*, 18(1), 1-25. <https://doi.org/10.1186/s12915-020-00820-5>

Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations?. *Trends in ecology & evolution*, 29(4), 233-242. <https://doi.org/10.1016/j.tree.2014.02.003>

Robertson, H. M. (2019). Molecular evolution of the major arthropod chemoreceptor gene families. *Annual review of entomology*, 64, 227-242. <https://doi.org/10.1146/annurev-ento-020117-043322>

Robertson, H. M., Robertson, E. C., Walden, K. K., Enders, L. S., & Miller, N. J. (2019). The chemoreceptors and odorant binding proteins of the soybean and pea aphids. *Insect biochemistry and molecular biology*, 105, 69-78. <https://doi.org/10.1016/j.ibmb.2019.01.005>

Roderick, G. K. (1996). Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annual review of entomology*, 41(1), 325-352. <https://doi.org/10.1146/annurev.en.41.010196.001545>

Russell JA, Weldon S, Smith AH, Kim KL, Hu Y, Łukasik P, Doll S, Anastopoulos I, Novin M, Oliver KM (2013) Uncovering symbiont-driven genetic diversity across North American pea aphids. *Molecular Ecology*, 22, 2045–2059. <https://doi.org/10.1111/mec.12211>

Sánchez-Gracia, A., Vieira, F. G., & Rozas, J. (2009). Molecular evolution of the major chemosensory gene families in insects. *Heredity*, 103(3), 208-216. <https://doi.org/10.1038/hdy.2009.55>

Santamaria, M. E., Arnaiz, A., Gonzalez-Melendi, P., Martinez, M., & Diaz, I. (2018). Plant perception and short-term responses to phytophagous insects and mites. *International journal of molecular sciences*, 19(5), 1356. <https://doi.org/10.3390/ijms19051356>

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.*, 38, 595-619. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095646>

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11), 807-820. <https://doi.org/10.1038/nrg3522>

Sicard, D., Pennings, P. S., Grandclément, C., Acosta, J., Kaltz, O., & Shykoff, J. A. (2007). Specialization and local adaptation of a fungal parasite on two host plant species as revealed by two fitness traits. *Evolution*, 61(1), 27-41. <https://doi.org/10.1111/j.1558-5646.2007.00003.x>

Sinclair, F. H., Stone, G. N., Nicholls, J. A., Cavers, S., Gibbs, M., Butterill, P., ... & Schönrogge, K. (2015). Impacts of local adaptation of forest trees on associations with herbivorous insects: implications for adaptive forest management. *Evolutionary Applications*, 8(10), 972-987. <https://doi.org/10.1111/eva.12329>

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics*, 6(1), 1-11. <https://doi.org/10.1186/1471-2105-6-31>

Smadja, C., Shi, P., Butlin, R. K., & Robertson, H. M. (2009). Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Molecular biology and evolution*, 26(9), 2073-2086. <https://doi.org/10.1093/molbev/msp116>

Shih, P. Y., Sugio, A., & Simon, J. C. (2023). Molecular mechanisms underlying host

plant specificity in Aphids. *Annual Review of Entomology*, 68, 431-450.
<https://doi.org/10.1146/annurev-ento-120220-020526>

Simon, J. C., d'Alencon, E., Guy, E., Jacquin-Joly, E., Jaquier, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in functional genomics*, 14(6), 413-423.
<https://doi.org/10.1093/bfgp/elv015>

Sinclair, F. H., Stone, G. N., Nicholls, J. A., Cavers, S., Gibbs, M., Butterill, P., ... & Schönrogge, K. (2015). Impacts of local adaptation of forest trees on associations with herbivorous insects: implications for adaptive forest management. *Evolutionary Applications*, 8(10), 972-987. <https://doi.org/10.1111/eva.12329>

Stukenbrock, E. H., & McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annu. Rev. Phytopathol.*, 46, 75-100.
<https://doi.org/10.1146/annurev.phyto.010708.154114>

Stukenbrock, E. H., & Bataillon, T. (2012). A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems.
<https://doi.org/10.1371/journal.ppat.1002893>

Stephan, W. (2019). Selective sweeps. *Genetics*, 211(1), 5-13.
<https://doi.org/10.1534/genetics.118.301319>

Srinivasan, D. G., & Brisson, J. A. (2012). Aphids: a model for polyphenism and epigenetics. *Genetics Research International*, 2012.
<https://doi.org/10.1155/2012/431531>

International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS biology*, 8(2), e1000313.
<https://doi.org/10.1371/journal.pbio.1000313>

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12), 673-680.
<https://doi.org/10.1016/j.tree.2014.10.004>

Thorpe, P., Escudero-Martinez, C. M., Cock, P. J., Eves-van den Akker, S., & Bos, J. I. (2018). Shared transcriptional control and disparate gain and loss of aphid

parasitism genes. *Genome biology and evolution*, 10(10), 2716-2733. <https://doi.org/10.1093/gbe/evy183>

Tsuchida T, Koga R, Shibao H, Matsumoto T, Fukatsu T (2002) Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, *Acyrtosiphon pisum*. *Molecular Ecology*, 11, 2123–2135. <https://doi.org/10.1046/j.1365-294X.2002.01606.x>

Wadgyamar, S. M., DeMarche, M. L., Josephs, E. B., Sheth, S. N., & Anderson, J. T. (2022). Local adaptation: Causal agents of selection and adaptive trait divergence. *Annual Review of Ecology, Evolution, and Systematics*, 53, 87-111. <https://doi.org/10.1146/annurev-ecolsys-012722-035231>

Weigel, D., & Nordborg, M. (2015). Population genomics for understanding adaptation in wild plant species. *Annual review of genetics*, 49, 315-338. <https://doi.org/10.1146/annurev-genet-120213-092110>

Williams, I. S., & Dixon, A. F. (2007). Life cycles and polymorphism. *Aphids as crop pests*, 69-85.

Whiteman, N. K., & Pierce, N. E. (2008). Delicious poison: genetics of *Drosophila* host plant preference. *Trends in Ecology & Evolution*, 23(9), 473-478. <https://doi.org/10.1016/j.tree.2008.05.010>

Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87-100. <https://doi.org/10.1038/nrg.2016.133>

Wyss, E., Villiger, M., & Müller-Schärer, H. (1999). The potential of three native insect predators to control the rosy apple aphid, *Dysaphis plantaginea*. *BioControl*, 44, 171-182. <https://doi.org/10.1023/A:1009934214927>

Zepeda-Paulo F, Ortiz-Martínez S, Silva AX, Lavandero B (2018) Low bacterial community diversity in two introduced aphid pests revealed with 16S rRNA amplicon sequencing. *PeerJ*, 6, e4725. <https://doi.org/10.7717/peerj.4725>

Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: progress and

challenges. *International journal of biological sciences*, 3(7), 420.
<https://doi.org/10.7150/ijbs.3.420>

Züst, T., & Agrawal, A. A. (2016). Mechanisms and evolution of plant resistance to aphids. *Nature plants*, 2(1), 1-9. <https://doi.org/10.1038/nplants.2015.206>

Zytynska, S. E., Frantz, L., Hurst, B., Johnson, A., Preziosi, R. F., & Rowntree, J. K. (2014). Host-plant genotypic diversity and community genetic interactions mediate aphid spatial distribution. *Ecology and evolution*, 4(2), 121-131.
<https://doi.org/10.1002/ece3.916>

3. INTRODUCTION

Genomic and ecological bases of adaptation in phytophagous insects: Insight from the aphids

DIVERGENCE PROCESSES

Even since Darwin's time, evolutionary biologists have attempted to tackle the evolutionary processes that shaped genetically differentiated populations or species in response to the environment (Wolf and Ellegren, 2017). Elucidating these questions also allows us to understand current and future threats to biodiversity (Pereira et al., 2010) and their consequences for ecosystem health and services (Cardinale et al., 2012).

The process during which groups of individuals evolve different genes, allele frequencies, and ultimately traits, and that can lead to new populations or species, is called divergence. Divergence can result from demographic (drift and/or gene flow) and/or adaptive (selection) processes (Figure 1). Over the past 15 years, the flourishing of sequencing data and the theoretical work and tools in population and comparative genomics allowed investigations of the role of demography and selection in the divergence of populations and species in response to biotic or abiotic environments (Nosil and Schluter, 2011; Tiffin and Ross-Ibarra, 2014). In particular, the local adaptive divergence of populations (or local adaptation), which represents the initial phase of reproductive isolation and ecological speciation (Lenormand, 2012; Karvonen and Seehausen, 2012; Figure 2), has been used to evaluate the contribution of specific environmental factors to the adaptive divergence (Savolainen et al., 2013). Local adaptation results from the spatial environmental variation of selection, which leads to locally adapted genotypes with higher fitness in their local than in their foreign abiotic or/and biotic environments

(Kawecki and Ebert, 2004; Salvoilainen et al., 2007; Wadgyamar et al., 2022 Figure 2). Several transplant experiments have been performed looking for patterns of local adaptation, mainly to the abiotic environment and in plants (Leimu and Fischer, 2008; Kawecki and Ebert, 2004; Blanquart et al., 2013; Sinclair et al., 2015; Nooten and Hughes, 2017; Capblancq et al., 2023; Meek et al., 2023; Razaque et al., 2023). Fitness estimates of three ecotypes of *Andropogon gerardii* Vitman (dry, mesic, and wet ecotypes) in a transplant experiment across an environmental gradient from dry to wet showed patterns of local adaptation in this species. Local adaptation was represented by the higher performance of the ecotypes under local conditions, with each ecotype exhibiting higher canopy cover (Johnson et al., 2021). In *Arabidopsis* (*Arabidopsis thaliana* L.), genomic and phenotypic variations associated with local adaptation to climate have been demonstrated. A study in Italian and Swedish *Arabidopsis* populations detected higher fitness of local genotypes at both sites and *Delay Of Germination 1* (*DOG1*) genes as candidate loci for local adaptation (Postma et al., 2016). In addition, a study analyzing a subset of SNPs in *A. thaliana* found that 10% of candidate climate-related SNPs with functions related to local adaptation, including photosynthetic active radiation, precipitation in the wettest month, precipitation in the driest month, length of the growing season, and relative humidity (Hancock et al., 2011). Despite the high number of phenotypic and genomic studies to date (Savolainen et al., 2013), there is a need for additional knowledge about the genomic basis of local adaptation to both abiotic and biotic environments (Gianoli and Valladares, 2012) and beyond plants. Therefore, some long-standing questions on the ecology and genomics of adaptation remain: What genomic architecture underlies (local) adaptive divergence (number, locations, and effects of genomic regions)? What form do adaptive changes take (amino-acid substitution, gene duplications, gene gains, gene losses, positive and purifying selection involved)? How do abiotic or biotic factors contribute to (local) adaptive divergence? During my Ph.D., I investigated those questions

using plant parasites as my biological system, as they are superb models for studying the process of divergence, and I combined comparative genomics, population genomics, and experiments.

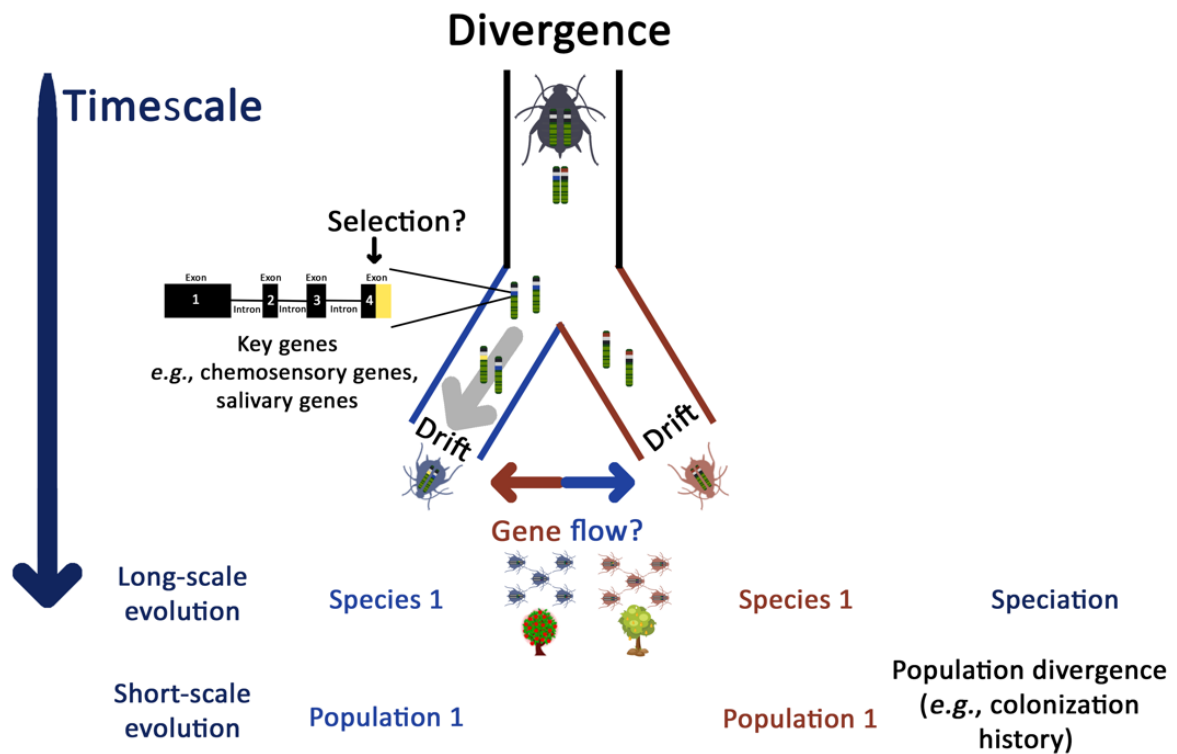


Figure 1. Evolutionary processes that can shape species and population divergence: insight from the aphids. Several evolutionary processes can shape the divergence of species and populations. Some genes related to response to biotic (e.g., chemosensory genes, salivary genes) or abiotic environments can undergo positive or purifying selective pressures that make those regions (genomic islands) less permeable to gene flow, ultimately leading to the divergence of species or populations. Drift and gene flow can favor or limit divergence. During my Ph.D., I investigated the role of selection, drift, and gene flow in the divergence of aphid species and populations.

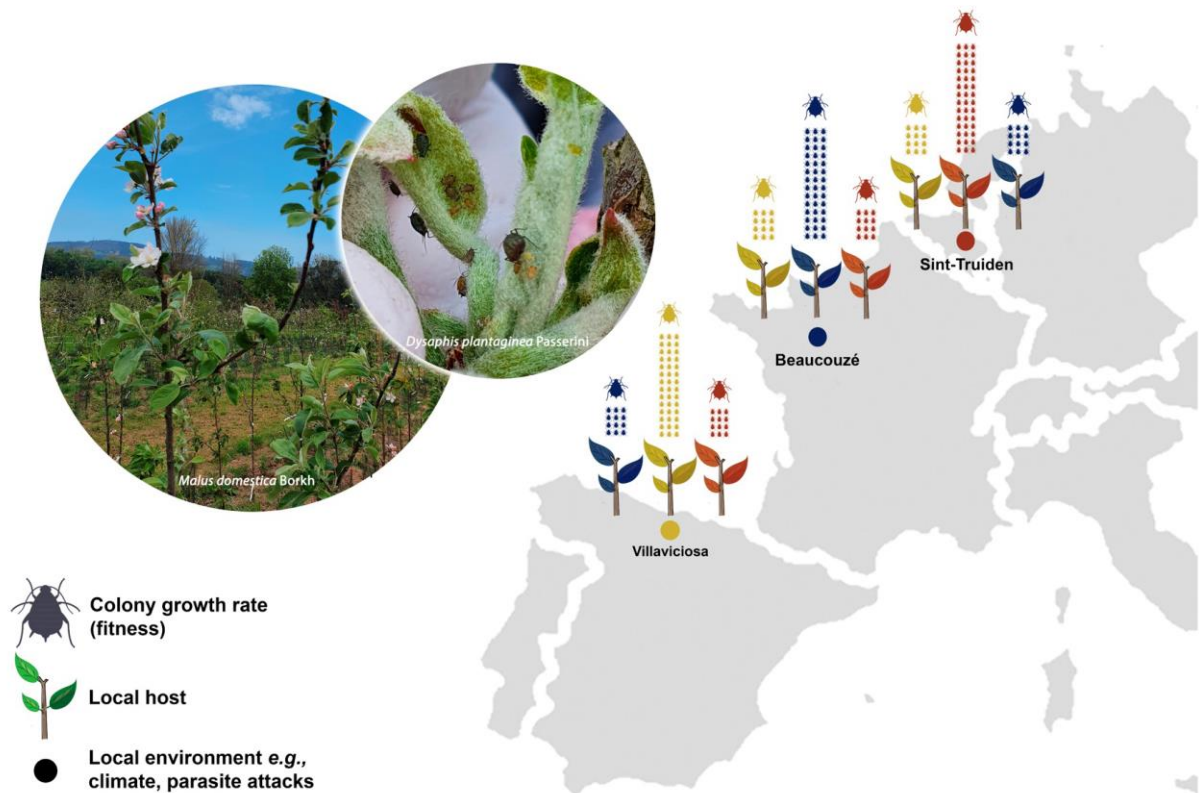


Figure 2. Expected patterns of local adaptation: An example from a plant parasite, the rosy apple aphid (*Dysaphis plantaginea* Passerini), a major pest of the cultivated apple (*Malus domestica* Borkh). Rosy apple aphid populations from three different origins (Sint-Truiden in Belgium, Beaucouzé in France, and Villaviciosa in Spain) are infested on local apple trees and placed under environmental conditions from the exact three origins. Patterns of local adaptation (the bar plots on the map) are when colony growth (considered here as fitness, *i.e.*, the reproductive success of a genotype) is higher for the local rosy apple aphid populations infested on their local apple tree host genotypes and/or local environmental conditions than those infested on foreign apple tree host genotypes and/or environmental conditions.

PLANT PARASITES: A GOOD STUDY MODEL TO UNDERSTAND PROCESSES OF DIVERGENCE

GLOBAL ADAPTIVE DIVERGENCE OF PLANT PARASITES TO THEIR HOST

Plant parasites are superb model systems for studying genomic and ecological bases of divergence. A parasite reduces its host ability to produce offspring (negative fitness effect) while spreading in the host population, positively impacting the fitness of the parasite (Méthot and Alizon, 2014). Reciprocally, the host counterattacks to overcome parasite attacks by evolving defense mechanisms. Therefore, the genomes of the interacting host-parasite constantly adapt such that each matches the change in the other (Van Valen, 1973). This reciprocal adaptation cycle led to the evolution of a parasite molecular arsenal, called effectors, to fight their host. Effectors help suppress the host defense and respond against the host immune system (Koeck et al., 2011; Yates and Michel, 2018). The evolution of specific effectors by gene duplications is a hypothesis that can explain the adaptive divergence of the parasites to their host by specialization (Möller and Stukenbrock, 2017; Vilcinskis, 2019). Indeed, duplications can generate new functions during species split that can be fixed by positive selection and contribute to species divergence (Ohno, 1970). After a long time, some copies of duplicated genes can be maintained by purifying selection (Yang, 2007; Cardoso-Moreira et al., 2016; Persi et al., 2016). Alternatively, the duplicated genes, here effector genes in case of parasite evolution, could present redundant functions, and a pseudogenization event might take place. In that case, duplicated genes become nonfunctional, resembling the original (in the structure), due to an accumulation of deleterious mutations in one of the gene copies (Assis and Bachtrog 2013). Thus, evaluating the selective pressures on the effector genes resulting from duplication can bring key information on evolutionary processes, the genomic bases underlying parasite divergence, and host-parasite co-evolution.

Transposable elements (TEs hereafter), DNA fragments that can replicate and multiply within genomes (Lisch, 2013), can be drivers of the parasite adaptation to their hosts. TEs enrichment and a high substitution rate close to effector genes have been detected in plant pathogens, such as *Zymoseptoria tritici* (Roberge ex Desm.) Quaedvl. & Crous (Hartman et al., 2017; review different fungi species, Plissonneau et al., 2018). TEs can, therefore, fuel the host-parasite arms race and parasite adaptation to their hosts. However, the role of TEs in driving the diversity and evolution of effector genes has been studied in plant pathogens (Duplessi et al., 2011; Plissonneau et al., 2017; Hartman et al., 2019; Rocafort et al., 2022). Narrow information exists about the characterization, numbers, nature, and mode of evolution of effector genes and the role of the TEs for other plant parasites, such as phytophagous insects (Stergiopoulos and de Wit, 2009; Simon et al., 2015; Torres et al., 2020; Seong and Krasileva, 2023).

LOCAL ADAPTATIVE DIVERGENCE IN PLANT PARASITES

Host-parasite coevolutionary dynamics are widely thought to lead to local adaptation of the parasite (Gandon and Van Zandt, 1998; Möller and Stukenbrock, 2017). Because the parasites have a shorter generation time than their hosts (Lajeunesse and Forbes, 2001), they evolve quicker, giving them the edge in the coevolutionary race (Stahl and Bishop, 2000; Zhan et al., 2002; Capaul and Ebert, 2003; Brockhurst et al., 2014; Sironi et al., 2015; Sheppard et al., 2018; Papkuo et al., 2019). However, local adaptation of the parasite is not always the rule; local adaptation depends on which partner is in the upper hand in this arms race and, therefore, on the respective generation times/evolutionary rates of the two partners (Kaltz & Shykoff, 1998). The spatial distribution of the host species is another factor that can impact the coevolutionary dynamics and, thus, local adaptation of

the parasite (Brown & Tellier, 2011; Gandon et al., 2008). Different pressures on the parasite arise from variable environments across space, resulting in spatial heterogeneity in parasite infectivity and host resistance (Caicedo and Schaal, 2004; Thrall and Burdon, 2002). The expansion of species range and the colonization of new habitats, for example, lead to the host encountering new parasites, potentially promoting coevolutionary dynamics at new loci or alleles not involved in coevolution in the original habitat (Lajeunesse and Forbes, 2002; Stam et al., 2019; Karasov et al., 2020; Markel et al., 2021, Cornille et al. 2022). Therefore, local adaptation of the parasites depends on the ecological context and the pressures imposed by the host (Savolainen et al., 2007; Savolainen et al., 2013).

Because of this theoretical framework, host-parasite interactions have been studied since a long time ago by evolutionary biologists to unravel the ecological and genomic bases and evolutionary processes that can lead to local adaptation (Badouin et al., 2017; Gladieux et al., 2014; Scott Nuismer & Sylvain Gandon, 2008). A handful of multi-species manipulative transplant experiments across environments have shown that the parasite affects the host's fitness (Greischar & Koskella, 2007; Scott Nuismer & Sylvain Gandon, 2008). Common gardens are also effective experimental designs to identify traits under local adaptation and the underlying ecological factors that have shaped that adaptation (Kaltz and Shykoff, 1998; Tiffin and Ross-Ibarra, 2014; Nooten and Hughes, 2017). However, it has been demonstrated that local adaptation of the parasite is not a universal phenomenon. Experimental approaches in the fungi *Microbotryum violaceum* (Pers.) G. Deml & Oberw. on its host *Silene latifolia* Poi. revealed a lower infection rate in sympatric groups (fungi populations infected in local populations) than in allopatric groups (fungi populations infected in foreign populations; Kaltz et al., 1999), suggesting maladaptation of the parasite, which has also been observed recently on its *Dianthus* L. host (Koupilová et al., 2021). As stated above, the spatial

distribution of the host species is crucial for coevolutionary dynamics (Gandon et al., 2008; Brown and Tellier, 2011), and different patterns of local adaptation could be seen across the host distribution, as local adaptation can depend on the host (biotic factors) and the local environments. However, multi-population manipulative transplant experiments across both host (*i.e.*, biotic) and abiotic environments reflecting a spatial heterogeneity, such as spatially varying climate, have never been done to disentangle the role of both biotic and abiotic factors in local adaptation of the parasite. Additionally, scans for signatures of selection or at candidate loci in parasite genomes have recently been performed to determine which partners are ahead in the game and which have achieved local adaptation. Genome scans applied to some parasites (Badouin et al., 2017), *e.g.*, tsetse flies (Attardo et al., 2019), malaria (Gupta et al., 2010), and phytophagous parasites (Zhu and Zhao, 2007; Bonin, 2008; Simon et al., 2015; Dalongeville et al., 2018), have revealed candidate genes under selection involved in local adaptation (Simon et al., 2015). Despite the current research contributions, the loci involved in local adaptation still need to be studied. Identifying loci under recent positive selection in populations is fundamental to elucidating adaptation mechanisms (Weigel & Nordborg, 2015).

Therefore, the direct links among local selective forces, adaptive phenotypic differentiation between populations, and the adaptive genomic variation underlying parasite local adaptation have yet to be fully elucidated (Gloss et al., 2016). Particularly in phytophagous insects (Nooten and Hughes, 2017), the ecological and genomics bases of the plant host divergence still need to be studied.

PHYTOPHAGOUS INSECT ADAPTIVE DIVERGENCE

GLOBAL ADAPTIVE DIVERGENCE OF PHYTOPHAGOUS INSECTS TO THEIR HOSTS: WHICH GENES INVOLVED AND MODE OF EVOLUTION? A COMPARATIVE GENOMICS VIEW

Preliminary works have paved the way to identifying the number and nature of genes involved in the adaptive divergence of phytophagous insects (Simon et al., 2015; Sih et al., 2023). Not only genes but also key genomic features like the transposable elements (TEs), which are repetitive regions of the genomes (Gilbert et al., 2021), have retained attention to understand the genomic bases of adaptive divergence in parasites. TEs have been proposed as a primary driver of genomic variation at key loci (*e.g.*, effectors genes), fueling the arms race between parasites and their host. However, the genomic architecture and the role of TEs in the evolution of key loci specifically involved in phytophagous insect divergence are still unidentified.

The annotation of chemosensory and salivary genes revealed their mode of evolution and potential role in adaptive divergence (Hogenhout and Bos, 2011; Simon et al., 2015; Züest and Agrawal, 2016; Nicolis et al., 2022; Shih et al., 2023). The interactions between plant chemical defense and phytophagous insects are processed by chemosensory receptors, including olfactory receptors (ORs) and gustatory receptors (GRs; Nimura and Nei et al., 2007; Nei et al., 2008; Sanchez-Garcia et al., 2009; Ramdya and Benton, 2010; Anderson et al., 2015; Benton, 2015). The ORs are seven-transmembrane domain proteins that participate in the semiochemical conversion into an electrical signal and are expressed in the olfactory neurons located within the olfactory sensilla (Clyne et al., 1999; Scott et al., 2001; Benton et al., 2006; Leal, 2013). The GRs are also a seven-transmembrane domain protein that responds to tastings (*e.g.*, pheromones, bitter substances)

and is situated within the gustatory sensilla (Chyb, 2004; Sparks et al., 2013; Jiang et al., 2015; Mang et al., 2016). In lepidopterans, multiple OR/GR gene gains were key in the adaptation of lepidopterans to their host plants (Engsontonia et al., 2014). The annotation of chemosensory gene families in several *Drosophila* Fallén genome assemblies revealed the rapid evolution of OR and GR gene families via lineage-specific duplications and loss events (Guo and Kim, 2007). Analysis of the correlation between genome size and duplication events of chemosensory gene families in lepidopterans and dipterans revealed positive correlations in *Drosophila* spp (Gardien et al., 2008). These genes would be explained by duplication events bringing novel functions in *Drosophila* spp, for instance, a new set of gustatory receptors genes to taste sugar (Kent and Robertson, 2009). On the other hand, negative correlations between genome size and chemosensory genes were detected in lepidopterans, suggesting that other genome features would explain the lepidopteran genome size and the current set of chemosensory genes is enough to ensure survival and reproduction (Engsontia et al., 2014). In hemipterans, recent and rapid expansion of the OR and GR genes were detected in the model species *Acyrtosiphon pisum* (the pea aphid), with some loci under positive selection (Smadja et al., 2009) and others under purifying selection. An additional analysis of 172 GR and OR genes showed high genetic differentiation among *A. pisum* biotypes (*i.e.*, individuals with some differences in their genomes and highly specialized to feed on different host legumes; Shih et al., 2023), resulting from host specialization (Smadja et al., 2012). This suggests that GR and OR can be responsible for host specialization, the first stage of the speciation process (process explained in Figure 1). Other important genes involved in host specialization are the salivary genes. For example, the salivary effectors of *Rhopalosiphum padi* (*RpC002*, *Rp1*), *Sitobion miscanthi* (*Sm9723*), and *Schizaphis graminum* (*Sg2204*) enhanced specific plant host susceptibility to aphid infestation (review in Muo et al., 2023). Therefore, host preferences can be a strong selective agent leading to the

differentiation of specific genes, which can head to speciation by adaptive divergence by host specialization. Genetic drift can also drive the evolution of genes in phytophagous insects. For instance, OR and GR in *Drosophila* Fallén spp are supposed to evolve by drift (Hallem et al., 2006; McBride, 2007; Sanchez-Gracia et al., 2009; Joseph and Carlson, 2016) and lead to divergence but not by adaptation. However, the balance between selection and drift in the evolution of genes involved in the divergence of phytophagous to their host still needs to be understood beyond model species such as butterflies, drosophila, and *A. pisum*. The time-consuming manual annotation and curation explain limited information about the OR and GR evolution beyond model insect species due to the high divergence among these genes. Moreover, the relationship between TEs and chemosensory gene evolution is an attractive topic because TEs are abundant in insect genomes (McCullers and Steiniger, 2017; Gilbert et al., 2021), and the TE content and type can vary among the genomes of insect species (Gilbert et al., 2021). However, the role of TEs in the evolution of the chemosensory genes (OR/GR) in phytophagous insects is not yet studied.

ON THE DEMOGRAPHY AND COLONIZATION HISTORY OF PHYTOPHAGOUS INSECTS: A PRELIMINARY AND MANDATORY INFORMATION FOR INVESTIGATING LOCAL ADAPTIVE DIVERGENCE

Before investigating the local adaptation of the target species, the knowledge of its population structure, diversity, and demographic history (divergence time, migration rates among populations, and effective population size) is essential to avoid false positives while testing for local adaptation (Li et al., 2012; Tiffin and Ross-Ibarra, 2014; Hoban et al., 2016). Studying population structure, genetic diversity, and demographic history is also fundamental to understanding evolutionary processes underlying rapid colonization and the range of geographic expansion (Estoup & Guillemaud, 2010; Lawson Handley et al., 2011; Turcotte et

al., 2017; Fraimout et al., 2017). The understanding of the routes of colonization of phytophagous insects can also generate information to protect crops against future outbreaks and for pest management programs (Posada & Crandall, 2001; Bertorelle et al., 2010; Bloomquist et al., 2010; Roux & Pannell, 2015; Hickerson et al., 2010). Demographic models have been tested using Approximate Bayesian computation methods (ABC) to understand the species history by comparing alternative demographic models and estimating their associated parameters (divergence time, migration rate, effective population size; Csilléry et al., 2010; Csilléry et al., 2012; Estoup et al., 2018; Raynal et al., 2019). The ABC approaches have been utilized to elucidate the evolutionary history identifying source population, colonization route, and geographic range expansion, like in *Daktulosphaira vitifoliae* Fitch (Rispe et al., 2020). Recently, ABC has been recently combined with machine learning, such as random forest and ABC approach, called ABC-RF (Estoup et al. 2018; Raynal et al. 2019), which gave a higher power to reconstruct the colonization history than traditional ABC methods (Csilléry et al., 2010; Csilléry et al., 2012). For example, ABC-RF approaches were utilized to reconstruct the colonization history of African arid-adapted locus species, *Schitocerca gregaria* Forsskål, in Africa through a major migration event by the last glacial climatic episodes (Chapuis et al., 2020). Also, the invasion history of *D. suzukii* proposes a route from Japan to Hawaii, Hawaii, and Southeast China to North America, and finally, an invasion to Europe from northeast China (Fraimout et al., 2017).

Characterizing the microbiota of the phytophagous insects can also bring information on the colonization history of their host. Indeed, phytophagous insects often host endosymbiotic bacteria that help them survive (Moran et al., 2008). Detecting variations in the microbiota composition of phytophagous insects along their colonization routes could suggest a potential role of the microbiota in the colonization of their phytophagous hosts. In the aphid populations of

Aphis craccivora Koch, or cowpea aphid possess a mix of facultative symbionts, and these bacteria composition can evolve rapidly in response to the environment imposed by the different host plants (Lenhart and White, 2020). In the pea aphid, different bacteria communities are associated with geographical variation (Tsuchida et al. 2002; Russell et al. 2013; Zepeda-Paulo et al. 2018; Leclair et al. 2021).

So far, the power of ABC-RF approaches and the characterization of the diversity of the aphid endosymbiotic bacterial community still need to be applied to other non-model phytophagous species. Results from these analyses bring essential information on the colonization history of the phytophagous insect and the study of genes involved in local adaptation.

LOCAL ADAPTIVE DIVERGENCE IN PHYTOPHAGOUS INSECTS

The genomic and ecological bases of local adaptation have been much more studied in plant pathogens than in phytophagous insects (Stukenbrock and Bataillon, 2012). For example, the finding of loci of avirulence effectors (*Avr*) related to host infection in *Zymoseptoria tritici* (Roberge ex Desm.) Quaedvl. & Crous (Croll and McDonald, 2017; Hartman et al., 2018), the diversification of effectors in the fungal wheat pathogen *Phaeosphaeria nodorum* (E.Müll.) Hedjar (Richards et al., 2019). In addition, 17% of the genome was under selective sweeps in the genome of *M. lychnidis-dioicae*, contrasting to 1% in the *Microbotryium silenes-dioicae* genome; these sweeps were in candidate effectors genes involved in parasite adaptation to its host (Badouin et al., 2017). In the crop pest *Plutella xylostella* (or diamondback moth), 3,648 SNPs were associated with specific climate variables (Chen et al., 2021) and, therefore, candidates for local adaptation to the abiotic

environment. In hemipteran, genes related to cathepsin B-like proteins with proteolytic activity to face the plant response in the midgut and saliva in the pea aphid were detected under positive selection (Review in Simon et al., 2015). In orthopters, 242 unique SNPs under selection related to latitude, mean annual temperature, and body size were detected in the grasshopper crop pest *Phaulacridium vittatum* Sjöstedt (Yadav et al., 2019). In the butterflies (or heliconia), genome scans detected a circadian clock gene (*Clk*) and two targets of pesticides in the oriental moth or *Grapholita molesta* Busck from Sichuan compared to the other two populations in Yunnan in China (Cao et al., 2022). Despite those numerous studies, the genomic regions underlying the local adaptation of phytophagous insects to their environment are generally unidentified, if any, and the selective pressures driving their evolution as well. Yet, phytophagous insects are ideal study subjects in the rich theoretical framework for local adaptation in host-parasite interactions (Gloss et al., 2016; Greischar & Koskella, 2007).

THE APPLE-APHID INTERACTION IS AN OUTSTANDING MODEL FOR INVESTIGATING THE ECOLOGICAL AND GENOMIC BASES OF THE ADAPTIVE DIVERGENCE OF PARASITES.

Studies of the mechanisms of adaptive divergence in the context of domestication are particularly relevant to crop parasites, which display frequent host shifts involving adaptation to new hosts (Gladieux et al., 2015; Powell et al., 2013; Stuckenberg and McDonald, 2008). Such host shifts are often followed by rapid colonization by the parasite of new geographic regions. Has parasite rapid colonization of different environments led to local adaptive divergence in the host-parasite interaction? What are the genomic bases of local adaptation in this context? Aphids are widely used as model organisms in evolutionary biology in studies of population differentiation in sympatry (Simon et al., 2015). Aphids are unusual in that they display cyclic parthenogenesis during their annual life cycle. Laboratory

experiments have shown that abiotic conditions (*e.g.*, temperature, humidity, photoperiod) are crucial for passage between steps in the life cycle. Aphids also depend on their host: they display plasticity for several characters, but many aphids show specificity for a particular host species. This dependence on biotic and abiotic conditions and the recent identification of gene-for-gene interactions during infestation (Züst & Agrawal, 2016) makes aphids a fascinating and relevant model for evaluating the contribution of abiotic and biotic conditions to local host-parasite adaptation.

Moreover, aphids have small genomes and can be reared in the laboratory. They are, therefore, ideal organisms for experimental manipulation and investigations of the experimental genomics of local adaptation in host-parasite interactions. Aphids are under different pressures to exploit their host: they recognize the chemical produced by their specific host, later they encounter additional physical and chemical barriers that avoid aphid feeding, and the ecological pressures by the host spatial variation. The development of host specialization, host races, and local adaptation could follow these pressures.

Dysaphis plantaginea, the rosy apple aphid, is the most harmful aphid pest attacking cultivated apple trees, causing significant economic losses every year. In cultivated apple orchards, this aphid species occurs across temperate regions (Asia, North Africa, America, and Europe). The ecology and evolution of *D. plantaginea* have been little explored to date (Guillemaud et al., 2011; Miñarro et al., 2005; Pagliarani et al., 2016; Qubbaj et al., 2005; Wyss et al., 1999). *Dysaphis plantaginea* probably colonized its cultivated apple host in Europe rapidly and recently, about 1,500 years ago, when the Greeks brought the cultivated apple (*Malus domestica*) to Europe from Central Asia (Cornille et al., 2014). Photoperiod is a crucial determinant of hatching, host colonization, and the induction of winged

and sexual forms in the *D. plantaginea* life cycle (Figure 3). We have observed a variation in hatching across Europe, with later and earlier hatching in Northern and Southern Europe, respectively. Whether this phenotypic variation reflects a local adaptation of *D. plantaginea* populations to the abiotic or biotic conditions following its recent colonization of Europe remains unclear. Moreover, the cultivated apple displays considerable variation in sensitivity and resistance to *D. plantaginea* attacks (Miñarro et al., 2005; Pagliarani et al., 2016), and *D. plantaginea* is very rarely seen on its wild European relative, *Malus sylvestris* L. (Mill.). Then, there is a need to evaluate the genomic architecture and the ecological bases behind the phenological variation and potentially local adaptation of the rosy apple aphid to its host, the apple tree, or its local environment.

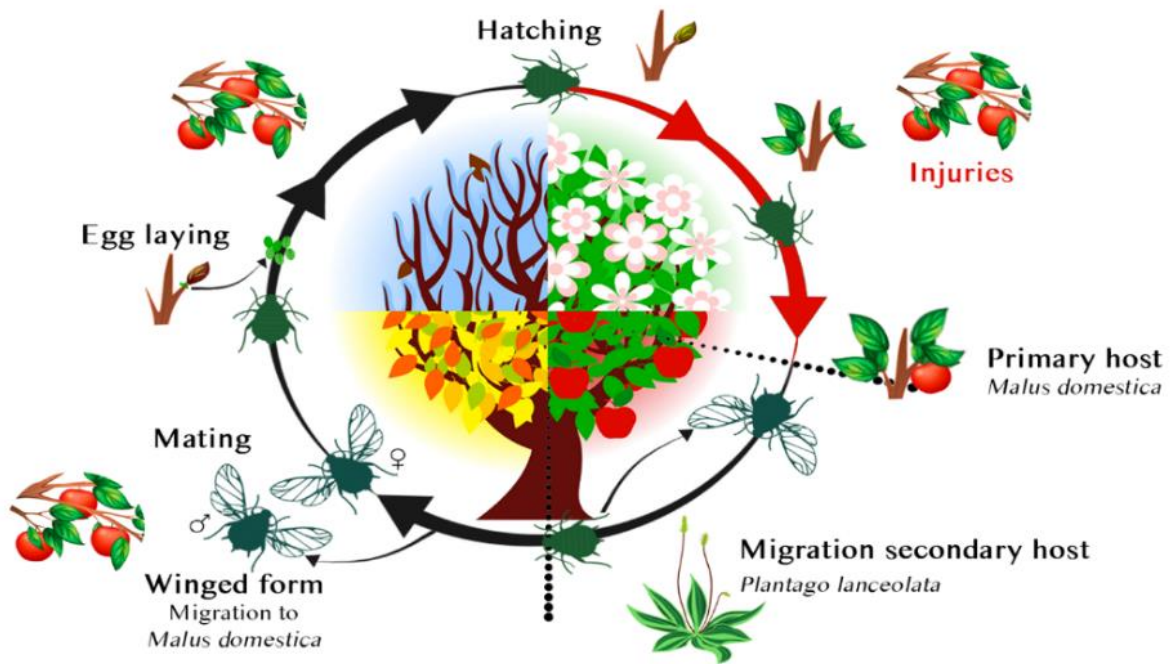


Figure 3. *Dysaphis plantaginea* life cycle. *Dysaphis plantaginea* has a cyclic parthenogenetic (or holocyclic) life cycle (Blommers et al., 2004; Guillemaud et al., 2011). Mating occurs on apples during the fall, and sexual females lay eggs that overwinter and then hatch at the beginning of spring. During late spring and early summer, after 3 to 6 parthenogenetic generations on apple, parthenogenetic females (virginiparae) produce winged morphs that migrate from the primary to the secondary host, *Plantago lanceolata* L., on which about 3 to 8 successive parthenogenetic generations occur. In late summer to fall, parthenogenetic female aphids (virginiparae) cyclically give birth to gynoparae (precursor forms of sexual females), followed by winged males. Both fly from the herbaceous secondary host plant, *P. lanceolata*, to the primary host, apple trees, where the gynoparae give birth to sexual females. Mating then occurs. Thus, due to the annual pattern of host alternation, two significant migration events occur in the biological cycle of *D. plantaginea* in the fall and spring.

OBJECTIVES OF THE PH.D. PROJECT: ON THE GENOMIC AND ECOLOGICAL BASES OF ADAPTIVE DIVERGENCE OF APHIDS IN THE CONTEXT OF DOMESTICATION

In this thesis dissertation, I aimed to unravel the genomic and ecological bases of adaptive divergence in the rosy apple aphid. I divided this thesis into four chapters presented in an article format, followed by a discussion of the main findings and a conclusion of the thesis project and perspectives of my work.

In the introduction section above, I presented the available information on the ecological and genomic bases of divergence in plant parasites, specifically in phytophagous insects. This section also introduced the biological model I worked on for four years, the rosy apple aphid or *Dysaphis plantaginea*, which is ideal for getting key information on phytophagous insect divergence. In Chapter 1, I investigated the adaptive divergence in aphids using comparative genomics. In Chapter 2, I inferred the colonization history of the rosy apple aphid using population genetics and microsatellite markers. In Chapter 3, I used population genomics and SNPs to understand the genomic bases of local adaptation in the rosy apple aphid. In Chapter 4, I tested for local adaptation using ecological data collected in a vast transplant and cross-infestation experiment.

Chapter 1: Chemosensory gene and transposable element evolutionary dynamics in aphids

This chapter aims to study the evolution of chemosensory genes during aphid evolution, as well as the role of transposable elements, which could have driven the evolution of chemosensory genes. We used comparative genomics approaches to confront different aphid species and one aphid-like species with various ecological preferences. Manual annotation of chemosensory genes was performed to get information about the genomic architecture of the host adaptation

in aphids. This manuscript will be submitted to the Molecular Biology and Evolution journal, for which I am the first author.

Chapter 2: Colonization history of the rosy apple aphid

Knowledge of the population structure and colonization history is crucial to investigate local adaptation. I, therefore, studied the colonization history of the rosy apple aphid using newly developed SSR markers. We detected recent and rapid colonization of cultivated apples (*Malus domestica*) by *D. plantaginea* in about 600 aphid samples representing 52 geographic sites. Weakly differentiated populations and substantial gene flow among rosy apple aphid populations were also detected. This work is published in the Peers in Community in Evolutionary Biology journal, supporting the free and transparent preprint pre-review Peers in Community in movement (<https://doi.org/10.24072/pcjournal.26>), in which I am the first author.

Chapter 3: Genomic bases of local adaptation in the rosy apple aphid (*Dysaphis plantaginea*) to its host, the cultivated apple (*Malus domestica*)

This section presents signatures of local adaptation in the genomes of the rosy apple aphid, which was investigated using SNP markers. The second manuscript, in which I am the first author, is under development and will be published in the fall of 2024.

Chapter 4: Experimental test for local adaptation of the rosy apple aphid (*Dysaphis plantaginea*) during its recent rapid colonization on its cultivated apple host (*Malus domestica*) in Europe

This chapter represents the experimental approach for seeking the ecological bases of adaptation of the rosy apple aphid to its local host or environment.

This chapter presents a common garden experiment repeated in Europe (Belgium, France, and Spain) to test whether different rosy apple aphid populations from Belgium, France, and Spain show differential fitness. Each common garden comprised a panel of wild and cultivated apple genotypes from Belgium, France, and Spain and previously reported tolerant and susceptible apple genotypes (Figure 2). The methodology of this chapter has been submitted as a pre-registration paper in the Peers Community in Ecology journal (https://forge-mia.inra.fr/amandine.cornille/local_adaptation_dp). The manuscript is progressing and will be published in the PCI Ecology journal, for which I will be the first author.

3.1 ANNEXES AND EXTRA ACTIVITIES

Olvera-Vazquez, S., G., ... Cornille, A., 2023. Participation in organizing the Young Scientist ESEB Workshop, the Experimental and Theoretical Approaches to Coevolution.

Chen X., ..., Olvera-Vazquez S. G., ..., Cornille A. 2023. Ecological and evolutionary drivers of phenotypic and genetic variation in the European crabapple [*Malus sylvestris* (L.) Mill.], a wild relative of the cultivated apple. *Annals of Botany*. <https://doi.org/10.1093/aob/mcad061>.

Cornille A., and Olvera-Vazquez, S. G., 2021. Une colonisation récente du puceron cendré du pommier dans les vergers en Europe, au Maghreb et en Amérique du Nord. *Passion Entomologie*. <https://passion-entomologie.fr/pommier-puceron-cendre/> (Science outreach).

3.2 STUDENT SUPERVISION

2023. Lilian Brouillard Master 1. Université Paris Cité. « Réponses des pommiers cultivés et sauvages au changement climatique et aux ravageurs ».

2022. Imane Elfar Master student 2. Université Évry. « Role of transposable elements in thirteen aphid species' evolution and colonization processes (Hemiptera: Aphididae) ».

2021. Ramon Gonzalez Master student 1. Universidad de Oviedo. «Practicum II Apple tree measurements and aphid cross infestation second part ».

2021. Cristina de Los Reyes Ramos Master student 2. Universidad de Oviedo. «Practicum II Apple tree measurements and aphid cross infestation ».

3.3 REFERENCES

Antonovics, J., Hood, M., & Partain, J. (2002). The ecology and genetics of a host shift: *Microbotryum* as a model system. *the american naturalist*, 160(S4), S40-S53. <https://doi.org/10.1086/342143>

Attardo, G. M., Abd-Alla, A. M., Acosta-Serrano, A., Allen, J. E., Bateta, R., Benoit, J. B., ... & Aksoy, S. (2019). Comparative genomic analysis of six *Glossina* genomes, vectors of African trypanosomes. *Genome biology*, 20(1), 1-31. <https://doi.org/10.1186/s13059-019-1768-2>

Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguileta, G., Snirc, A., ... & Giraud, T. (2017). Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Molecular Ecology*, 26(7), 2041-2062. <https://doi.org/10.1111/mec.13976>

Bazzicalupo, A. (2022). Local adaptation in fungi. *FEMS Microbiology Reviews*, 46(6), fuac026. <https://doi.org/10.1093/femsre/fuac026>

Benton, R. (2015). Multigene family evolution: perspectives from insect chemoreceptors. *Trends in ecology & evolution*, 30(10), 590-600. <https://doi.org/10.1016/j.tree.2015.07.009>

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular ecology*, 19, 2609–2625. <https://doi.org/10.1111/j.1365-294X.2010.04690.x>

Bonin, A. (2008). Population genomics: a new generation of genome scans to bridge the gap with functional genomics. <https://doi.org/10.1111/j.1365-294X.2008.03854.x>

Bonte, D., De Roissart, A., Vandegehuchte, M. L., Ballhorn, D. J., Van Leeuwen, T., & de la Pena, E. (2010). Local adaptation of aboveground herbivores towards plant phenotypes induced by soil biota. *PLoS One*, 5(6), e11174. <https://doi.org/10.1371/journal.pone.0011174>

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology letters*, 16(9), 1195-1205. <https://doi.org/10.1111/ele.12150>

Blommers, L. H. M., Helsen, H. H. M., & Vaal, F. W. N. M. (2004). Life history data of the rosy apple aphid *Dysaphis plantaginea* (Pass.) (Homopt., Aphididae) on plantain and as migrant to apple. *Journal of Pest Science*, 77, 155-163. <https://doi.org/10.1007/s10340-004-0046-5>

Bloomquist EW, Lemey P, Suchard MA (2010) Three roads diverged? Routes to phylogeographic inference. *Trends in Ecology & Evolution*, 25, 626-632. <https://doi.org/10.1016/j.tree.2010.08.010>

Bouallègue, M., Filée, J., Kharrat, I., Mezghani-Khemakhem, M., Rouault, J. D., Makni, M., & Capy, P. (2017). Diversity and evolution of mariner-like elements in aphid genomes. *BMC genomics*, 18(1), 1-12. <https://doi.org/10.1186/s12864-017-3856-6>

Braendle, C., Davis, G. K., Brisson, J. A., & Stern, D. L. (2006). Wing dimorphism in aphids. *Heredity*, 97(3), 192-199. <https://doi.org/10.1038/sj.hdy.6800863>

Brown, J. K., & Tellier, A. (2011). Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual review of phytopathology*, 49, 345-367. <https://doi.org/10.1146/annurev-phyto-072910-095301>

Caicedo, A. L., & Schaal, B. A. (2004). Heterogeneous evolutionary processes affect R gene diversity in natural populations of *Solanum pimpinellifolium*. *Proceedings of the National Academy of Sciences*, 101(50), 17444-17449. <https://doi.org/10.1073/pnas.0407899101>

Cao, L. J., Song, W., Chen, J. C., Fan, X. L., Hoffmann, A. A., & Wei, S. J. (2022). Population genomic signatures of the oriental fruit moth related to the Pleistocene climates. *Communications Biology*, 5(1), 142. <https://doi.org/10.1038/s42003-022-03097-2>

Capblancq, T., Lachmuth, S., Fitzpatrick, M. C., & Keller, S. R. (2023). From common gardens to candidate genes: exploring local adaptation to climate in red spruce. *New Phytologist*, 237(5), 1590-1605. <https://doi.org/10.1111/nph.18465>

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., ... & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486(7401), 59-67. <https://doi.org/10.1038/nature11148>

Cardoso-Moreira, M., Arguello, J. R., Gottipati, S., Harshman, L. G., Grenier, J. K., & Clark, A. G. (2016). Evidence for the fixation of gene duplications by positive selection in *Drosophila*. *Genome research*, 26(6), 787-798. <https://doi.org/10.1101/gr.199323.115>

Chapuis M-P, Raynal L, Plantamp C, Meynard CN, Blondin L, Marin J-M, Estoup A (2020) A young age of subspecific divergence in the desert locust inferred by ABC random forest. *Molecular Ecology*, 29, 4542–4558. <https://doi.org/10.1111/mec.15663>

Chen, Y., Liu, Z., Régnière, J., Vasseur, L., Lin, J., Huang, S., ... & You, S. (2021). Large-scale genome-wide study reveals climate adaptive variability in a cosmopolitan pest. *Nature Communications*, 12(1), 7206. <https://doi.org/10.1038/s41467-021-27510-2>

Chyb, S. (2004). *Drosophila* gustatory receptors: from gene identification to functional expression. *Journal of insect physiology*, 50(6), 469-477. <https://doi.org/10.1016/j.jinsphys.2004.03.012>

Clyne, P. J., Warr, C. G., Freeman, M. R., Lessing, D., Kim, J., & Carlson, J. R. (1999). A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron*, 22(2), 327-338. [https://doi.org/10.1016/S0896-6273\(00\)81093-4](https://doi.org/10.1016/S0896-6273(00)81093-4)

Cornille, A., Giraud, T., Smulders, M. J., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, 30(2), 57-65. <https://doi.org/10.1016/j.tig.2013.10.002>

Cornille, A., Ebert, D., Stukenbrock, E., de la Vega, R. C. R., Tiffin, P., Croll, D., & Tellier, A. (2022). Unraveling coevolutionary dynamics using ecological genomics. *Trends in Genetics*. <https://doi.org/10.1016/j.tig.2022.05.008>

Croll, D., & McDonald, B. A. (2017). The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Molecular ecology*, 26(7), 2027-2040. <https://doi.org/10.1111/mec.13870>

Csilléry K, Blum MG, Gaggiotti OE, François O (2010) Approximate Bayesian computation (ABC) in practice. *Trends in ecology & evolution*, 25, 410–418. <https://doi.org/10.1016/j.tree.2010.04.001>

Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3, 475–479. <https://doi.org/10.1111/j.2041-210X.2011.00179.x>

Dalongeville, A., Benestan, L., Mouillot, D., Lobreaux, S., & Manel, S. (2018). Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). *BMC genomics*, 19, 1-13. <https://doi.org/10.1186/s12864-018-4579-z>

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, 19, 4113–4130. <https://doi.org/10.1111/j.1365-294X.2010.04773.x>

Estoup A, Raynal L, Verdu P, Marin J-M (2018) Model choice using Approximate Bayesian Computation and Random Forests: analyses based on model grouping to make inferences about the genetic history of Pygmy human populations. *Journal de la Société Française de Statistique*, 159, 167–190. http://www.numdam.org/item/JSFS_2018__159_3_167_0/

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481–501. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173414>

Fraimout A, Debat V, Fellous S, Hufbauer RA, Foucaud J, Pudlo P, Marin J-M, Price DK, Cattell J, Chen X, Deprá M, François Duyck P, Guedot C, Kenis M, Kimura MT, Loeb G, Loiseau A, Martinez-Sañudo I, Pascual M, Polihronakis Richmond M, Shearer P, Singh N, Tamura K, Xuéreb A, Zhang J, Estoup A (2017) Deciphering the Routes of invasion of *Drosophila suzukii* by Means of ABC Random Forest. *Molecular Biology and Evolution*, 34, 980–996. <https://doi.org/10.1093/molbev/msx050>

Gandon, S., & Van Zandt, P. A. (1998). Local adaptation and host–parasite interactions. *Trends in Ecology & Evolution*, 13(6), 214-216. [https://doi.org/10.1016/S0169-5347\(98\)01358-5](https://doi.org/10.1016/S0169-5347(98)01358-5)

Gandon, S., Buckling, A., Decaestecker, E., & Day, T. (2008). Host–parasite coevolution and patterns of adaptation across time and space. *Journal of evolutionary biology*, 21(6), 1861-1866. <https://doi.org/10.1111/j.1420-9101.2008.01598.x>

Garrido, E., Andraca-Gómez, G., & Fornoni, J. (2012). Local adaptation: simultaneously considering herbivores and their host plants. *New Phytologist*, 193(2), 445-453. <https://doi.org/10.1111/j.1469-8137.2011.03923.x>

Gianoli, E., & Valladares, F. (2012). Studying phenotypic plasticity: the advantages of a broad approach. *Biological Journal of the Linnean Society*, 105(1), 1-7. <https://doi.org/10.1111/j.1095-8312.2011.01793.x>

Gienapp, P., Teplitsky, C., Alho, J. S., Mills, J. A., & Merilä, J. (2008). Climate change and evolution: disentangling environmental and genetic responses. *Molecular ecology*, 17(1), 167-178 <https://doi.org/10.1111/j.1365-294X.2007.03413.x>

Gilbert, C., Peccoud, J., & Cordaux, R. (2021). Transposable elements and the evolution of insects. *Annual Review of Entomology*, 66(1), 355-372. <https://doi.org/10.1146/annurev-ento-070720-074650>

Gloss, A. D., Groen, S. C., & Whiteman, N. K. (2016). A genomic perspective on the generation and maintenance of genetic diversity in herbivorous insects. *Annual review of ecology, evolution, and systematics*, 47, 165-187. <https://doi.org/10.1146/annurev-ecolsys-121415-032220>

Greischar, M. A., & Koskella, B. (2007). A synthesis of experimental work on parasite local adaptation. *Ecology letters*, 10(5), 418-434. <https://doi.org/10.1111/j.1461-0248.2007.01028.x>

Guillemaud, T., Blin, A., Simon, S., Morel, K., & Franck, P. (2011). Weak spatial and temporal population genetic structure in the rosy apple aphid, *Dysaphis plantaginea*, in French apple orchards. *Plos one*, 6(6), e21263. <https://doi.org/10.1371/journal.pone.0021263>

Gupta, B., Awasthi, G., & Das, A. (2010). Malaria parasite genome scan: insights into antimalarial resistance. *Parasitology research*, 107, 495-499. <https://doi.org/10.1007/s00436-010-1917-8>

Guillemaud, T., Blin, A., Simon, S., Morel, K., & Franck, P. (2011). Weak spatial and temporal population genetic structure in the rosy apple aphid, *Dysaphis plantaginea*, in French apple orchards. *Plos one*, 6(6), e21263. <https://doi.org/10.1371/journal.pone.0021263>

Hallem, E. A., Dahanukar, A., & Carlson, J. R. (2006). Insect odor and taste receptors. *Annu. Rev. Entomol.*, 51, 113-135 <https://doi.org/10.1146/annurev.ento.51.051705.113646>

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., ... & Bergelson, J. (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, 334(6052), 83-86. <https://doi.org/10.1126/science.1209244>

Hartmann, F. E., McDonald, B. A., & Croll, D. (2018). Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Molecular ecology*, 27(12), 2725-2741. <https://doi.org/10.1111/mec.14711>

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010) Phylogeography's past, present, and future: 10 years after. *Molecular phylogenetics and evolution*, 54, 291-301. <https://doi.org/10.1016/j.ympev.2009.09.016>

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., ... & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4), 379-397. <https://doi.org/10.1086/688018>

Hogehout, S. A., & Bos, J. I. (2011). Effector proteins that modulate plant-insect interactions. *Current opinion in plant biology*, 14(4), 422-428. <https://doi.org/10.1016/j.pbi.2011.05.003>

Jaenike, J. (1990). Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics*, 21(1), 243-273. <https://doi.org/10.1146/annurev.es.21.110190.001331>

Jiang, X. J., Ning, C., Guo, H., Jia, Y. Y., Huang, L. Q., Qu, M. J., & Wang, C. Z. (2015). A gustatory receptor tuned to D-fructose in antennal sensilla chaetica of *Helicoverpa armigera*. *Insect Biochemistry and Molecular Biology*, 60, 39-46. <https://doi.org/10.1016/j.ibmb.2015.03.002>

Johnson, L. C., Galliard, M. B., Alsdurf, J. D., Maricle, B. R., Baer, S. G., Bello, N. M., ... & Smith, A. B. (2022). Reciprocal transplant gardens as gold standard to detect local adaptation in grassland species: New opportunities moving into the 21st century. *Journal of Ecology*, 110(5), 1054-1071. [file:///Reciprocal transplant gardens as gold standard to detect local adaptation in grassland species/ New opportunities moving into the 21st century](file:///Reciprocal%20transplant%20gardens%20as%20gold%20standard%20to%20detect%20local%20adaptation%20in%20grassland%20species/%20New%20opportunities%20moving%20into%20the%2021st%20century)

Kaloshian, I., & Walling, L. L. (2016). Hemipteran and dipteran pests: effectors and plant host immune regulators. *Journal of Integrative Plant Biology*, 58(4), 350-361. <https://doi.org/10.1111/jipb.12438>

Kaltz, O., & Shykoff, J. A. (1998). Local adaptation in host-parasite systems. *Heredity*, 81(4), 361-370. <https://doi.org/10.1046/j.1365-2540.1998.00435.x>

Kaltz, O., Gandon, S., Michalakis, Y., & Shykoff, J. A. (1999). Local maladaptation in the anther-smut fungus *Microbotryum violaceum* to its host plant *Silene latifolia*: evidence from a cross-inoculation experiment. *Evolution*, 53(2), 395-407. <https://doi.org/10.1111/j.1558-5646.1999.tb03775.x>

Karasov, T. L., Shirsekar, G., Schwab, R., & Weigel, D. (2020). What natural variation can teach us about resistance durability. *Current Opinion in Plant Biology*, 56, 89-98. <https://doi.org/10.1016/j.pbi.2020.04.010>

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology letters*, 7(12), 1225-1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>

Keller, I., & Seehausen, O. (2012). Thermal adaptation and ecological speciation. *Molecular ecology*, 21(4), 782-799. <https://doi.org/10.1111/j.1365-294X.2011.05397.x>

Kent, L. B., & Robertson, H. M. (2009). Evolution of the sugar receptors in insects. *BMC Evolutionary Biology*, 9, 1-20. <https://doi.org/10.1186/1471-2148-9-41>

Koskella, B., Giraud, T., & Hood, M. E. (2006). Pathogen relatedness affects the prevalence of within-host competition. *The American Naturalist*, 168(1), 121-126. <https://doi.org/10.1086/505770>

Koupilová, K., Koubek, T., Cornille, A., & Janovský, Z. (2021). Local maladaptation of the anther-smut fungus parasitizing *Dianthus carthusianorum*. *European Journal of Plant Pathology*, 160(2), 365-374. <https://doi.org/10.1007/s10658-021-02249-0>

Lawson Handley L-J, Estoup A, Evans DM, Thomas CE, Lombaert E, Facon B, Aebi A, Roy HE (2011) Ecological genetics of invasive alien species. *BioControl*, 56, 409–428. <https://doi.org/10.1007/s10526-011-9386-2>

Lajeunesse, M. J., & Forbes, M. R. (2002). Host range and local parasite adaptation. *Proceedings of the royal society of london. Series b: biological sciences*, 269(1492), 703-710. <https://doi.org/10.1098/rspb.2001.1943>

Leal, W. S. (2013). Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annual review of entomology*, 58, 373-391. <https://doi.org/10.1146/annurev-ento-120811-153635>

Leclair M, Buchard C, Mahéo F, Simon J-C, Outreman Y (2021) A Link Between Communities of Protective Endosymbionts and Parasitoids of the Pea Aphid Revealed

in Unmanipulated Agricultural Systems. *Frontiers in Ecology and Evolution*, 9. <https://doi.org/10.3389/fevo.2021.618331>

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PloS one*, 3(12), e4010. <https://doi.org/10.1371/journal.pone.0004010>

Lenhart PA, White JA (2020) Endosymbionts facilitate rapid evolution in a polyphagous herbivore. *Journal of Evolutionary Biology*, 33, 1507–1511. <https://doi.org/10.1111/jeb.13697>

Li, J., Li, H., Jakobsson, M., Li, S. E. N., SjöDin, P. E. R., & Lascoux, M. (2012). Joint analysis of demography and selection in population genetics: where do we stand and where could we go?. *Molecular ecology*, 21(1), 28-44. <https://doi.org/10.1111/j.1365-294X.2011.05308.x>

Lin, R., Yang, M., & Yao, B. (2022). The phylogenetic and evolutionary analyses of detoxification gene families in Aphidinae species. *PloS one*, 17(2), e0263462. <https://doi.org/10.1371/journal.pone.0263462>

Lisch, D. (2013). How important are transposons for plant evolution?. *Nature Reviews Genetics*, 14(1), 49-61. <https://doi.org/10.1038/nrg3374>

Mang, D., Shu, M., Endo, H., Yoshizawa, Y., Nagata, S., Kikuta, S., & Sato, R. (2016). Expression of a sugar clade gustatory receptor, BmGr6, in the oral sensory organs, midgut, and central nervous system of larvae of the silkworm *Bombyx mori*. *Insect biochemistry and molecular biology*, 70, 85-98. <https://doi.org/10.1016/j.ibmb.2015.12.008>

Martel, A., Ruiz-Bedoya, T., Breit-McNally, C., Laflamme, B., Desveaux, D., & Guttman, D. S. (2021). The ETS-ETI cycle: evolutionary processes and metapopulation dynamics driving the diversification of pathogen effectors and host immune factors. *Current Opinion in Plant Biology*, 62, 102011. <https://doi.org/10.1016/j.pbi.2021.102011>

Meek, M. H., Beever, E. A., Barbosa, S., Fitzpatrick, S. W., Fletcher, N. K., Mittan-Moreau, C. S., ... & Hellmann, J. J. (2023). Understanding local adaptation to prepare populations for climate change. *BioScience*, 73(1), 36-47. <https://doi.org/10.1093/biosci/biac101>

Merilä, J., & Hendry, A. P. (2014). Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary applications*, 7(1), 1-14. <https://doi.org/10.1111/eva.12137>

Méthot, P. O., & Alizon, S. (2014). What is a pathogen? Toward a process view of host-parasite interactions. *Virulence*, 5(8), 775-785. <https://doi.org/10.4161/21505594.2014.960726>

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, 15(12), 756-771. <https://doi.org/10.1038/nrmicro.2017.76>

Moran, N. A., McCutcheon, J. P., & Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annual review of genetics*, 42, 165-190. <https://doi.org/10.1146/annurev.genet.41.110306.130119>

Nicolis, V. F., Burger, N. F. V., & Botha, A. M. (2022). Whole-body transcriptome mining for candidate effectors from *Diuraphis noxia*. *BMC genomics*, 23(1), 493. <https://doi.org/10.1186/s12864-022-08712-4>

Niimura, Y., & Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PloS one*, 2(8), e708. <https://doi.org/10.1371/journal.pone.0000708>

Nooten, S. S., & Hughes, L. (2017). The power of the transplant: direct assessment of climate change impacts. *Climatic Change*, 144, 237-255. <https://doi.org/10.1007/s10584-017-2037-6>

Nouhaud, P., Gautier, M., Gouin, A., Jaquiéry, J., Peccoud, J., Legeai, F., ... & Simon, J. C. (2018). Identifying genomic hotspots of differentiation and candidate genes involved in the adaptive divergence of pea aphid host races. *Molecular Ecology*, 27(16), 3287-3300. <https://doi.org/10.1111/mec.14799>

Oakley, C. G., Schemske, D. W., McKay, J. K., & Ågren, J. (2023). Ecological genetics of local adaptation in *Arabidopsis*: An 8-year field experiment. *Molecular Ecology*. <https://doi.org/10.1111/mec.17045>

Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, New York).

Olvera-Vazquez, S. G., Remoué, C., Venon, A., Rousselet, A., Grandcolas, O., Azrine, M., ... & Cornille, A. (2021). Large-scale geography survey provides insights into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa. *Peer Community Journal*, 1. <https://doi.org/10.24072/pcjournal.26>

Peccoud, Jean, & Simon, Jean-Christophe. (2010). The pea aphid complex as a model of ecological speciation. *Ecological Entomology*, 35, 119-130. <https://doi.org/10.1111/j.1365-2311.2009.01147.x>

Pereira, H. M., Leadley, P. W., Proença, V., Alkemade, R., Scharlemann, J. P., Fernandez-Manjarrés, J. F., ... & Walpole, M. (2010). Scenarios for global biodiversity in the 21st century. *Science*, 330(6010), 1496-1501. <https://doi.org/10.1126/science.1196624>

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in ecology & evolution*, 16, 37-45. [https://doi.org/10.1016/S0169-5347\(00\)02026-7](https://doi.org/10.1016/S0169-5347(00)02026-7)

Postma, F. M., & Ågren, J. (2016). Early life stages contribute strongly to local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 113(27), 7590-7595. <https://doi.org/10.1073/pnas.1606303113>

Qubbaj, T., Reineke, A., & Zebitz, C. P. W. (2005). Molecular interactions between rosy apple aphids, *Dysaphis plantaginea*, and resistant and susceptible cultivars of its primary host *Malus domestica*. *Entomologia Experimentalis et Applicata*, 115(1), 145-152. <https://doi.org/10.1111/j.1570-7458.2005.00255.x>

Raynal L, Marin J-M, Pudlo P, Ribatet M, Robert CP, Estoup A (2019) ABC random forests for Bayesian parameter inference. *Bioinformatics*, 35, 1720-1728. <https://doi.org/10.1093/bioinformatics/bty867>

Razzaque, S., Heckman, R. W., & Juenger, T. E. (2023). Seed size variation impacts local adaptation and life-history strategies in a perennial grass. *Proceedings of the Royal Society B*, 290(1998), 20222460. <https://doi.org/10.1098/rspb.2022.2460>

Richards, J. K., Stukenbrock, E. H., Carpenter, J., Liu, Z., Cowger, C., Faris, J. D., & Friesen, T. L. (2019). Local adaptation drives the diversification of effectors in the fungal wheat pathogen *Parastagonospora nodorum* in the United States. *PLoS Genetics*, 15(10), e1008223. <https://doi.org/10.1371/journal.pgen.1008223>

Rispe C, Legeai F, Nabity PD, Fernández R, Arora AK, Baa-Puyoulet P, Banfill CR, Bao L, Barberà M, Bouallègue M, Bretaudeau A, Brisson JA, Calevro F, Capy P, Catrice O, Chertemps T, Couture C, Delière L, Douglas AE, Dufault-Thompson K, Escuer P, Feng H, Forneck A, Gabaldón T, Guigó R, Hilliou F, Hinojosa-Alvarez S, Hsiao Y, Hudaverdian S, Jacquin-Joly E, James EB, Johnston S, Joubard B, Le Goff G, Le Trionnaire G, Librado P, Liu S, Lombaert E, Lu H, Maibèche M, Makni M, Marcet-

Houben M, Martínez-Torres D, Meslin C, Montagné N, Moran NA, Papura D, Parisot N, Rahbé Y, Lopes MR, Ripoll-Cladellas A, Robin S, Roques C, Roux P, Rozas J, Sánchez-Gracia A, Sánchez-Herrero JF, Santesmasses D, Scatoni I, Serre R-F, Tang M, Tian W, Umina PA, van Munster M, Vincent-Monégat C, Wemmer J, Wilson ACC, Zhang Y, Zhao C, Zhao J, Zhao S, Zhou X, Delmotte F, Tagu D (2020) The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC Biology*, 18, 90. <https://doi.org/10.1186/s12915-020-00820-5>

Roux C, Pannell JR (2015) Inferring the mode of origin of polyploid species from next-generation sequence data. *Molecular Ecology*, 24, 1047–1059. <https://doi.org/10.1111/mec.13078>

Russell JA, Weldon S, Smith AH, Kim KL, Hu Y, Łukasik P, Doll S, Anastopoulos I, Novin M, Oliver KM (2013) Uncovering symbiont-driven genetic diversity across North American pea aphids. *Molecular Ecology*, 22, 2045–2059. <https://doi.org/10.1111/mec.12211>

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.*, 38, 595-619. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095646>

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11), 807-820. <https://doi.org/10.1038/nrg3522>

Scott, K., Brady Jr, R., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C., & Axel, R. (2001). A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell*, 104(5), 661-673. [https://doi.org/10.1016/S0092-8674\(01\)00263-X](https://doi.org/10.1016/S0092-8674(01)00263-X)

Shih, P. Y., Sugio, A., & Simon, J. C. (2023). Molecular mechanisms underlying host plant specificity in Aphids. *Annual Review of Entomology*, 68, 431-450. <https://doi.org/10.1146/annurev-ento-120220-020526>

Sicard, D., Pennings, P. S., Grandclément, C., Acosta, J., Kaltz, O., & Shykoff, J. A. (2007). Specialization and local adaptation of a fungal parasite on two host plant species as revealed by two fitness traits. *Evolution*, 61(1), 27-41. <https://doi.org/10.1111/j.1558-5646.2007.00003.x>

Sinclair, F. H., Stone, G. N., Nicholls, J. A., Cavers, S., Gibbs, M., Butterill, P., ... & Schönrogge, K. (2015). Impacts of local adaptation of forest trees on associations

with herbivorous insects: implications for adaptive forest management. *Evolutionary Applications*, 8(10), 972-987. <https://doi.org/10.1111/eva.12329>

Simon, J. C., Rispe, C., & Sunnucks, P. (2002). Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, 17(1), 34-39. [https://doi.org/10.1016/S0169-5347\(01\)02331-X](https://doi.org/10.1016/S0169-5347(01)02331-X)

Simon, J. C., Stoeckel, S., & Tagu, D. (2010). Evolutionary and functional insights into reproductive strategies of aphids. *Comptes Rendus Biologies*, 333(6-7), 488-496. <https://doi.org/10.1016/j.crv.2010.03.003>

Simon, J. C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquiere, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in functional genomics*, 14(6), 413-423. <https://doi.org/10.1093/bfgp/elv015>

Smadja, C., Shi, P., Butlin, R. K., & Robertson, H. M. (2009). Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Molecular biology and evolution*, 26(9), 2073-2086. <https://doi.org/10.1093/molbev/msp116>

Smadja, C. M., Canbäck, B., Vitalis, R., Gautier, M., Ferrari, J., Zhou, J. J., & Butlin, R. K. (2012). Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution*, 66(9), 2723-2738. <https://doi.org/10.1111/j.1558-5646.2012.01612.x>

Stam, R., Silva-Arias, G. A., & Tellier, A. (2019). Subsets of NLR genes show differential signatures of adaptation during colonization of new habitats. *New Phytologist*, 224(1), 367-379. <https://doi.org/10.1111/nph.16017>

Stukenbrock, E. H., & Bataillon, T. (2012). A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. <https://doi.org/10.1371/journal.ppat.1002893>

Sork, V. L., Stowe, K. A., & Hochwender, C. (1993). Evidence for local adaptation in closely adjacent subpopulations of northern red oak (*Quercus rubra* L.) expressed as resistance to leaf herbivores. *The American Naturalist*, 142(6), 928-936. <https://doi.org/10.1086/285581>

Sparks, J. T., Vinyard, B. T., & Dickens, J. C. (2013). Gustatory receptor expression in the labella and tarsi of *Aedes aegypti*. *Insect biochemistry and molecular biology*, 43(12), 1161-1171. <https://doi.org/10.1016/j.ibmb.2013.10.005>

Srinivasan, D. G., & Brisson, J. A. (2012). Aphids: a model for polyphenism and epigenetics. *Genetics Research International*, 2012. <https://doi.org/10.1155/2012/431531>

Thrall, P. H., & Burdon, J. J. (2002). Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal. *Plant Pathology*, 51(2), 169-184. <https://doi.org/10.1046/j.1365-3059.2002.00683.x>

Tack, A. J., & Roslin, T. (2010). Overrun by the neighbors: landscape context affects strength and sign of local adaptation. *Ecology*, 91(8), 2253-2260. <https://doi.org/10.1890/09-0080.1>

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12), 673-680. <https://doi.org/10.1016/j.tree.2014.10.004>

Tsuchida T, Koga R, Shibao H, Matsumoto T, Fukatsu T (2002) Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, *Acyrtosiphon pisum*. *Molecular Ecology*, 11, 2123–2135. <https://doi.org/10.1046/j.1365-294X.2002.01606.x>

Turcotte MM, Araki H, Karp DS, Poveda K, Whitehead SR (2017) The eco-evolutionary impacts of domestication and agricultural practices on wild species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372, 20160033. <https://doi.org/10.1098/rstb.2016.0033>

Velasco-Cuervo, S. M., Galindo-González, L., & Toro-Perea, N. (2023). An omics evolutionary perspective on phytophagous insect–host plant interactions in *Anastrepha obliqua*: a review. *Entomologia Experimentalis et Applicata*, 171(1), 2-16. <https://doi.org/10.1111/eea.13241>

Wadgyamar, S. M., DeMarche, M. L., Josephs, E. B., Sheth, S. N., & Anderson, J. T. (2022). Local adaptation: Causal agents of selection and adaptive trait divergence. *Annual Review of Ecology, Evolution, and Systematics*, 53, 87-111. <https://doi.org/10.1146/annurev-ecolsys-012722-035231>

Wang, H., Shi, S., & Hua, W. (2023). Advances of herbivore-secreted elicitors and effectors in plant-insect interactions. *Frontiers in Plant Science*, 14, 1176048. <https://doi.org/10.3389/fpls.2023.1176048>

Weigel, D., & Nordborg, M. (2015). Population genomics for understanding adaptation in wild plant species. *Annual review of genetics*, 49, 315-338. <https://doi.org/10.1146/annurev-genet-120213-092110>

Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87-100. <https://doi.org/10.1038/nrg.2016.133>

Zepeda-Paulo F, Ortiz-Martínez S, Silva AX, Lavandero B (2018) Low bacterial community diversity in two introduced aphid pests revealed with 16S rRNA amplicon sequencing. *PeerJ*, 6, e4725. <https://doi.org/10.7717/peerj.4725>

Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *International journal of biological sciences*, 3(7), 420. <https://doi.org/10.7150/ijbs.3.420>

Zovi, D., Stastny, M., Battisti, A., & Larsson, S. (2008). Ecological costs on local adaptation of an insect herbivore imposed by host plants and enemies. *Ecology*, 89(5), 1388-1398. <https://doi.org/10.1890/07-0883.1>

Züst, T., & Agrawal, A. A. (2016). Mechanisms and evolution of plant resistance to aphids. *Nature plants*, 2(1), 1-9. <https://doi.org/10.1038/nplants.2015.206>

4. OUTLINE CHAPTER 1 : CHEMOSENSORY GENES AND TRANSPOSABLE ELEMENTS EVOLUTIONARY DYNAMICS IN APHIDS

Identifying the number and nature of genes involved in the adaptive divergence of phytophagous insects to their host still needs to be studied. Plants offer insects an extensive range of habitats, food resources, mating, and oviposition sites (Lee and Lee, 2020) and are proposed to drive the diversification of phytophagous insects (Abrahamson, 2008). So far, the identification of genes involved in the adaptation of the phytophagous insects to their host and their mode of evolution by drift (Barton, 1996; White et al., 2020; Lynch, 2023) and/or natural selection (Bustamantes et al., 2005; Fay and Wittkop, 2008; Karlsson et al., 2014; Arnab et al., 2023; Culver et al., 2023) are topics of intense investigations. There is also an increasing interest in understanding the role of repetitive regions of the genome, or transposable elements (TEs hereafter; Gilbert et al., 2010), in driving the genomic variation fueling the adaptation of parasites, including phytophagous insects, to their host (Fouché et al., 2022; Meslin et al., 2022; Oggenfuss and Croll, 2023).

Chemosensation (olfaction and taste) is crucial in the interactions between phytophagous insects and their plant hosts. Phytophagous insects distinguish the chemicals the plants produce by activating specific proteins that multigene families encode (Simon et al., 2015; Robertson, 2019). Chemosensory genes are highly diversified multigene families (Santamaria et al., 2018) proposed as crucial in the adaptation of phytophagous insects to their host (Kambere and Lane, 2007; De Bruyne and Baker, 2008; Brand et al., 2015; Rinker et al., 2019). The olfactory receptors (ORs) and gustatory receptors (GRs) are significant chemoreceptor super-families in insects (Sanchez-Gracia et al., 2009; Robertson et al., 2019). The OR and GR genes were proposed to evolve following a birth-and-death model, where gene

duplications represent the births while the deaths are characterized by pseudogenization and deletion events (Nimura and Nei et al., 2008; Sanchez-Garcia et al., 2009; Ramdya and Benton, 2010; Anderson et al., 2015; Benton, 2015). The studies have detected that the OR and GR genes have a high evolutionary rate, leading to a vast number of gene numbers with evidence of expansion or contraction, depending on the chemosensory genes and insect species (Halem et al., 2006; McBride, 2007; Joseph and Carlson, 2016; Sanchez-Gracia et al., 2009). The elements of the genomic architecture (*i.e.*, transposable elements or TEs) and the selective pressures (*i.e.*, positive and negative selection) are among the significant drivers of the evolutionary dynamics of these chemosensory genes (Whiteman and Pierce, 2008; Meslin et al., 2022). For instance, high densities of TEs were related to duplication events in different insect species (Gilbert, Peccoud, and Cordaux, 2021). In addition, selection signatures were reported in the chemosensory genes (including ORs and GRs) of the pea aphid, which can be implicated in the divergence of pea aphid host races (Smadja et al., 2009; Meslin et al., 2022). Nevertheless, the evolution of OR and GR genes and the implication of TEs in phytophagous insect host preference remains to be studied.

Aphids are phytophagous insect groups, including several crop pests (Stern 2008), and are used as models for exploring the evolutionary processes involved in adaptation to the host and its associated genomic basis (Peccoud and Simon, 2010; Simon et al., 2015; *i.e.*, TEs; Bouallégue et al., 2017; Ahmad et al., 2021). Aphids have high intraspecific genetic and phenotypic diversity (Loxdale, 2008; Zytynska et al., 2013; Leclair et al., 2016) and exhibit diverse life cycles presenting polyphenism, sexual and asexual reproduction (Srinivasan and Brisson., 2012). Aphids have a range of host preferences. For instance, aphids would have a particular host species or several host species from a plant family to complete their

life cycle (Dixon, 1977; Williams and Dixon., 2007). Aphids rely on their chemosensory system to perceive chemicals, obtain environmental information, and detect their host (Robertson et al., 2019). These characteristics are essential for aphid adaptation to different environments and hosts. Current available aphid genome evolutionary genomics studies revealed that most species underwent extensive gene duplication through aphid diversification (The International Aphid Genomics Consortium, 2010; Mathers et al., 2017; Thorpe et al., 2018; Fernandez et al., 2020; Julca et al., 2020; Biello et al., 2021). Moreover, rapid gene evolution and genes enriched in essential functions for host selection and adaptation (e.g., chemosensory receptors genes and genes related to detoxification) are suggested as key for aphid specialization to their host (Smadja et al., 2009; Puinean et al., 2010; Nicholson et al., 2015; Quan et al., 2019).

In Chapter 1, we reconstructed the evolutionary history of two major chemosensory multigene families in a comprehensive sample of aphid species with various host breath to determine the mode of evolution of OR and GR genes and gain insight into the selective forces that drive their evolution in major phytophagous insects. We also characterized TE diversity and dynamics in the aphid genomes to understand the role of TEs in the evolution of chemosensory genes in aphids. We selected 11 aphid species (see details in the manuscript) whose genome was available. We also added the high-quality genome assembly genome version 3 of the rosy apple aphid (*Dysaphis plantaginea*) assembled in the team. We also added one aphid-like species of the Phylloxeridae family, *Daktulosphaira vitifoliae* Fitch, which was used as an outgroup for our analyses.

We manually annotated 521 ORs and 399 GRs among nine aphid species using three different software Exonerate (Slater and Birney, 2005), the InsectOR pipeline (Karpe et al., 2021), and Scipio (Keller et al., 2008). We did not identify new

OR and GR genes in *A. pisum*, *A. glycines*, and *D. vitifoliae*; we, therefore, used the already published chemosensory genes 147 ORs and 181 GRs (Robertson et al., 2019; Risper et al., 2020). These genes followed the birth-and-death model, evolved in tandem repeats, and under purifying selection. Only a few genes showed signatures of positive selection, including gustatory receptor gene group GR02 related to fructose detection (60% of the sites under positive selection). We also showed that TEs did not have a role in fueling GR and OR gene evolution. However, we found species-specific TE activities, with recent activity of transposable elements on aphids infesting fruit trees.

We obtained information on the molecular evolution of fundamental genes involved in the aphid adaptation to their hosts. For the rosy apple aphid, we got results suggesting a recent divergence of the species, very recent TE activity, and TIR TEs related to olfactory receptor genes. This comparative genomics analysis allowed us to better understand the evolutionary processes at play during the evolution of aphids at a large evolutionary scale. This chapter is in the process of submission to the journal *Molecular Biology and Evolution*.

OUTLINE REFERENCES

Abrahamson, W. G. (2008). *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects*. Univ of California Press.

Ahmad, A., Wallau, G. L., & Ren, Z. (2021). Characterization of Mariner transposons in seven species of Rhus gall aphids. *Scientific reports*, 11(1), 1-11. <https://doi.org/10.1038/s41598-021-95843-5>

Alexa, A., & Rahnenführer, J. (2009). Gene set enrichment analysis with topGO. *Bioconductor Improv*, 27, 1-26. <https://bioconductor.statistik.tu-dortmund.de/packages/3.3/bioc/vignettes/topGO/inst/doc/topGO.pdf>

Almeida-Silva, F., Zhao, T., Ullrich, K. K., Schranz, M. E., & Van de Peer, Y. (2023). syntenet: an R/Bioconductor package for the inference and analysis of synteny networks. *Bioinformatics*, 39(1), btac806. <https://doi.org/10.1093/bioinformatics/btac806>

Andersson, M. N., Löfstedt, C., & Newcomb, R. D. (2015). Insect olfaction and the evolution of receptor tuning. *Frontiers in Ecology and Evolution*, 3, 53. <https://doi.org/10.3389/fevo.2015.00053>

Arnab, S. P., Amin, M. R., & DeGiorgio, M. (2023). Uncovering footprints of natural selection through spectral analysis of genomic summary statistics. *Molecular Biology and Evolution*, msad157. <https://doi.org/10.1093/molbev/msad157>

Barton, N. H. (1996). Natural selection and random genetic drift as causes of evolution on islands. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1341), 785-795. <https://doi.org/10.1098/rstb.1996.0073>

Benton, R., Sachse, S., Michnick, S. W., & Vosshall, L. B. (2006). Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. *PLoS biology*, 4(2), e20. <https://doi.org/10.1371/journal.pbio.0040020>

Benton, R. (2015). Multigene family evolution: perspectives from insect chemoreceptors. *Trends in ecology & evolution*, 30(10), 590-600. <https://doi.org/10.1016/j.tree.2015.07.009>

Biello, R., Singh, A., Godfrey, C. J., Fernández, F. F., Mugford, S. T., Powell, G., ... & Mathers, T. C. (2021). A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Molecular ecology resources*, 21(1), 316-326. <https://doi.org/10.1111/1755-0998.13258>

Bouallègue, M., Filée, J., Kharrat, I., Mezghani-Khemakhem, M., Rouault, J. D., Makni, M., & Capy, P. (2017). Diversity and evolution of mariner-like elements in aphid genomes. *BMC genomics*, *18*(1), 1-12. <https://doi.org/10.1186/s12864-017-3856-6>

Brand, P., Ramírez, S. R., Leese, F., Quezada-Euan, J. J. G., Tollrian, R., & Eltz, T. (2015). Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). *BMC evolutionary biology*, *15*(1), 1-16. <https://doi.org/10.1186/s12862-015-0451-9>

Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Todd Hubisz, M., Glanowski, S., ... & Clark, A. G. (2005). Natural selection on protein-coding genes in the human genome. *Nature*, *437*(7062), 1153-1157. <https://doi.org/10.1038/nature04240>

Chen, W., Shakir, S., Bigham, M., Richter, A., Fei, Z., & Jander, G. (2019). Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). *Gigascience*, *8*(4), giz033. <https://doi.org/10.1093/gigascience/giz033>

Chyb, S. (2004). Drosophila gustatory receptors: from gene identification to functional expression. *Journal of insect physiology*, *50*(6), 469-477. <https://doi.org/10.1016/j.jinsphys.2004.03.012>

Clyne, P. J., Warr, C. G., Freeman, M. R., Lessing, D., Kim, J., & Carlson, J. R. (1999). A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron*, *22*(2), 327-338. [https://doi.org/10.1016/S0896-6273\(00\)81093-4](https://doi.org/10.1016/S0896-6273(00)81093-4)

Culver, D. C., Kowalko, J. E., & Pipan, T. (2023). Natural selection versus neutral mutation in the evolution of subterranean life: A false dichotomy?. *Frontiers in Ecology and Evolution*, *11*, 1080503. <https://doi.org/10.3389/fevo.2023.1080503>

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., ... & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2), giab008. <https://doi.org/10.1093/gigascience/giab008>

De Bruyne, M., & Baker, T. C. (2008). Odor detection in insects: volatile codes. *Journal of chemical ecology*, *34*(7), 882-897. <https://doi.org/10.1007/s10886-008-9485-4>

Dixon, A. F. G. (1977). Aphid ecology: life cycles, polymorphism, and population regulation. *Annual Review of Ecology and Systematics*, 329-353. <https://www.jstor.org/stable/2096732>

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, 20(1), 1-14. <https://doi.org/10.1186/s13059-019-1832-y>

Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic acids research*, 30(7), 1575-1584. <https://doi.org/10.1093/nar/30.7.1575>

Fay, J. C., & Wittkopp, P. J. (2008). Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*, 100(2), 191-199. <https://doi.org/10.1038/sj.hdy.6801000>

Fernández, R., Marcet-Houben, M., Legeai, F., Richard, G., Robin, S., Wucher, V., ... & Tagu, D. (2020). Selection following gene duplication shapes recent genome evolution in the pea aphid *Acyrtosiphon pisum*. *Molecular biology and evolution*, 37(9), 2601-2615. <https://doi.org/10.1093/molbev/msaa110>

Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PloS one*, 6(1), e16526. <https://doi.org/10.1371/journal.pone.0016526>

Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit. 2019. "RepeatModeler2: Automated Genomic Discovery of Transposable Element Families." Preprint. Genomics. <https://doi.org/10.1101/856591>

Fouché, S., Oggenfuss, U., Chanclud, E., & Croll, D. (2022). A devil's bargain with transposable elements in plant pathogens. *Trends in Genetics*, 38(3), 222-230. <https://doi.org/10.1016/j.tig.2021.08.005>

Gilbert, C., Peccoud, J., & Cordaux, R. (2021). Transposable elements and the evolution of insects. *Annual Review of Entomology*, 66(1), 355-372. <https://doi.org/10.1146/annurev-ento-070720-074650>

Hallem, E. A., Dahanukar, A., & Carlson, J. R. (2006). Insect odor and taste receptors. *Annu. Rev. Entomol.*, 51, 113-135 <https://doi.org/10.1146/annurev.ento.51.051705.113646>

International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS biology*, 8(2), e1000313. <https://doi.org/10.1371/journal.pbio.1000313>

Jiang, X. J., Ning, C., Guo, H., Jia, Y. Y., Huang, L. Q., Qu, M. J., & Wang, C. Z. (2015). A gustatory receptor tuned to D-fructose in antennal sensilla chaetica of *Helicoverpa armigera*. *Insect Biochemistry and Molecular Biology*, 60, 39-46. <https://doi.org/10.1016/j.ibmb.2015.03.002>

Jiang, X., Zhang, Q., Qin, Y., Yin, H., Zhang, S., Li, Q., ... & Chen, J. (2019). A chromosome-level draft genome of the grain aphid *Sitobion miscanthi*. *Gigascience*, 8(8), giz101. <https://doi.org/10.1093/gigascience/giz101>

Joseph, R. M., & Carlson, J. R. (2015). Drosophila chemoreceptors: a molecular interface between the chemical world and the brain. *Trends in Genetics*, 31(12), 683-695. <https://doi.org/10.1016/j.tig.2015.09.005>

Julca, I., Marcet-Houben, M., Cruz, F., Vargas-Chavez, C., Johnston, J. S., Gómez-Garrido, J., ... & Gabaldón, T. (2020). Phylogenomics identifies an ancestral burst of gene duplications predating the diversification of Aphidomorpha. *Molecular Biology and Evolution*, 37(3), 730-756. <https://doi.org/10.1093/molbev/msz261>

Kambere, M. B., & Lane, R. P. (2007). Co-regulation of a large and rapidly evolving repertoire of odorant receptor genes. *BMC neuroscience*, 8(3), 1-16. <https://doi.org/10.1186/1471-2202-8-S3-S2>

Karlsson, E. K., Kwiatkowski, D. P., & Sabeti, P. C. (2014). Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, 15(6), 379-393. <https://doi.org/10.1038/nrg3734>

Karpe, S. D., Tiwari, V., & Ramanathan, S. (2021). InsectOR—Webserver for sensitive identification of insect olfactory receptor genes from non-model genomes. *Plos one*, 16(1), e0245324. <https://doi.org/10.1371/journal.pone.0245324>

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, 20(4), 1160-1166. <https://doi.org/10.1093/bib/bbx108>

Keller, O., Odronitz, F., Stanke, M., Kollmar, M., & Waack, S. (2008). Scipio: using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC bioinformatics*, 9(1), 1-12. <https://doi.org/10.1186/1471-2105-9-278>

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, 16(2), 111-120. <https://doi.org/10.1007/BF01731581>

Kuraku, S., Zmasek, C. M., Nishimura, O., & Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic acids research*, 41(W1), W22-W28. <https://doi.org/10.1093/nar/gkt389>

Leal, W. S. (2013). Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annual review of entomology*, 58, 373-391. <https://doi.org/10.1146/annurev-ento-120811-153635>

Leclair, M., Pons, I., Mahéo, F., Morlière, S., Simon, J. C., & Outreman, Y. (2016). Diversity in symbiont consortia in the pea aphid complex is associated with large phenotypic variation in the insect host. *Evolutionary Ecology*, 30(5), 925-941. <https://doi.org/10.1007/s10682-016-9856-1>

Lee, S., & Lee, S. (2020). Multigene phylogeny uncovers oviposition-related evolutionary history of Cerambycinae (Coleoptera: Cerambycidae). *Molecular phylogenetics and evolution*, 145, 106707. <https://doi.org/10.1016/j.ympev.2019.106707>

Legeai, F., Shigenobu, S., Gauthier, J. P., Colbourne, J., Rispe, C., Collin, O., ... & Tagu, D. (2010). AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome. *Insect molecular biology*, 19, 5-12. <https://doi.org/10.1111/j.1365-2583.2009.00930.x>

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic acids research*, 49(W1), W293-W296. <https://doi.org/10.1093/nar/gkab301>

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493-496. <https://doi.org/10.1038/nature10231>

Loxdale, H. D. (2008). The nature and reality of the aphid clone: genetic variation, adaptation and evolution. *Agricultural and Forest Entomology*, 10(2), 81-90. <https://doi.org/10.1111/j.1461-9563.2008.00364.x>

Lynch, M. (2023). Mutation pressure, drift, and the pace of molecular coevolution. *Proceedings of the National Academy of Sciences*, 120(27), e2306741120. <https://doi.org/10.1073/pnas.2306741120>

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42, 847-862. doi: [10.3758/BRM.42.3.847](https://doi.org/10.3758/BRM.42.3.847)

Mang, D., Shu, M., Endo, H., Yoshizawa, Y., Nagata, S., Kikuta, S., & Sato, R. (2016). Expression of a sugar clade gustatory receptor, BmGr6, in the oral sensory organs, midgut, and central nervous system of larvae of the silkworm *Bombyx mori*. *Insect biochemistry and molecular biology*, 70, 85-98. <https://doi.org/10.1016/j.ibmb.2015.12.008>

Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., Baa-Puyoulet, P., ... & Hogenhout, S. A. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome biology*, 18(1), 1-20. <https://doi.org/10.1186/s13059-016-1145-3>

Mathers, T. C. (2020). Improved genome assembly and annotation of the soybean aphid (*Aphis glycines* Matsumura). *G3: Genes, Genomes, Genetics*, 10(3), 899-906. <https://doi.org/10.1534/g3.119.400954>

Mathers, T. C., Mugford, S. T., Hogenhout, S. A., & Tripathi, L. (2020). Genome sequence of the banana aphid, *Pentalonia nigronervosa* Coquerel (Hemiptera: Aphididae) and its symbionts. *G3: Genes, Genomes, Genetics*, 10(12), 4315-4321. <https://doi.org/10.1534/g3.120.401358>

Mathers, T. C., Wouters, R. H., Mugford, S. T., Swarbreck, D., Van Oosterhout, C., & Hogenhout, S. A. (2021). Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Molecular Biology and Evolution*, 38(3), 856-875. <https://doi.org/10.1093/molbev/msaa246>

McBride, C. S. (2007). Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proceedings of the National Academy of Sciences*, 104(12), 4996-5001. <https://doi.org/10.1073/pnas.0608424104>

Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518 (2020). <https://doi.org/10.1093/bioinformatics/btaa1022>

Meslin, Camille, Pauline Mainet, Nicolas Montagné, Stéphanie Robin, Fabrice Legeai, Anthony Bretaudeau, J Spencer Johnston, et al. 2022. "Spodoptera Littoralis Genome Mining Brings Insights on the Dynamic of Expansion of Gustatory Receptors in Polyphagous Noctuidae." *G3 Genes|Genomes|Genetics*. <https://doi.org/10.1093/g3journal/jkac131>

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37(5), 1530-1534. <https://doi.org/10.1093/molbev/msaa015>

Nicholson, S. J., Nickerson, M. L., Dean, M., Song, Y., Hoyt, P. R., Rhee, H., ... & Puterka, G. J. (2015). The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC genomics*, 16(1), 1-16. <https://doi.org/10.1186/s12864-015-1525-1>

Niimura, Y., & Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS one*, 2(8), e708. <https://doi.org/10.1371/journal.pone.0000708>

Oggenfuss, U., & Croll, D. (2023). Recent transposable element bursts are associated with the proximity to genes in a fungal plant pathogen. *PLoS Pathogens*, 19(2), e1011130. <https://doi.org/10.1371/journal.ppat.1011130>

Peccoud, Jean, & Simon, Jean-Christophe. (2010). The pea aphid complex as a model of ecological speciation. *Ecological Entomology*, 35, 119-130. <https://doi.org/10.1111/j.1365-2311.2009.01147.x>

Puinean, A. M., Foster, S. P., Oliphant, L., Denholm, I., Field, L. M., Millar, N. S., ... & Bass, C. (2010). Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLoS genetics*, 6(6), e1000999. <https://doi.org/10.1371/journal.pgen.1000999>

Quan, Q., Hu, X., Pan, B., Zeng, B., Wu, N., Fang, G., ... & Zhan, S. (2019). Draft genome of the cotton aphid *Aphis gossypii*. *Insect Biochemistry and Molecular Biology*, 105, 25-32. <https://doi.org/10.1016/j.ibmb.2018.12.007>

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., & Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS computational biology*, 1(2), e22. <https://doi.org/10.1371/journal.pcbi.0010022>

Ramdya, P., & Benton, R. (2010). Evolving olfactory systems on the fly. *Trends in Genetics*, 26(7), 307-316. <https://doi.org/10.1016/j.tig.2010.04.004>

Rinker, D. C., Specian, N. K., Zhao, S., & Gibbons, J. G. (2019). Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. *Proceedings of the National Academy of Sciences*, 116(27), 13446-13451. <https://doi.org/10.1073/pnas.1901093116>

Rispe, C., Legeai, F., Nabity, P. D., Fernández, R., Arora, A. K., Baa-Puyoulet, P., ... & Tagu, D. (2020). The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC biology*, 18(1), 1-25. <https://doi.org/10.1186/s12915-020-00820-5>

Robertson, H. M., Warr, C. G., & Carlson, J. R. (2003). Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 100(suppl_2), 14537-14542. <https://doi.org/10.1073/pnas.233584710>

Robertson, H. M. (2019). Molecular evolution of the major arthropod chemoreceptor gene families. *Annual review of entomology*, 64, 227-242. <https://doi.org/10.1146/annurev-ento-020117-043322>

Robertson, H. M., Robertson, E. C., Walden, K. K., Enders, L. S., & Miller, N. J. (2019). The chemoreceptors and odorant binding proteins of the soybean and pea aphids. *Insect biochemistry and molecular biology*, 105, 69-78. <https://doi.org/10.1016/j.ibmb.2019.01.005>

Sánchez-Gracia, A., Vieira, F. G., & Rozas, J. (2009). Molecular evolution of the major chemosensory gene families in insects. *Heredity*, 103(3), 208-216. <https://doi.org/10.1038/hdy.2009.55>

Santamaria, M. E., Arnaiz, A., Gonzalez-Melendi, P., Martinez, M., & Diaz, I. (2018). Plant perception and short-term responses to phytophagous insects and mites. *International journal of molecular sciences*, 19(5), 1356. <https://doi.org/10.3390/ijms19051356>

Scott, K., Brady Jr, R., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C., & Axel, R. (2001). A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell*, 104(5), 661-673. [https://doi.org/10.1016/S0092-8674\(01\)00263-X](https://doi.org/10.1016/S0092-8674(01)00263-X)

Sheffield, N. C., & Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*, 32(4), 587-589. <https://doi.org/10.1093/bioinformatics/btv612>

Simon, J. C., d'Alencon, E., Guy, E., Jacquin-Joly, E., Jaquier, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in functional genomics*, 14(6), 413-423. <https://doi.org/10.1093/bfgp/elv015>

Smadja, C., Shi, P., Butlin, R. K., & Robertson, H. M. (2009). Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Molecular biology and evolution*, 26(9), 2073-2086. <https://doi.org/10.1093/molbev/msp116>

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics*, 6(1), 1-11. <https://doi.org/10.1186/1471-2105-6-31>

Srinivasan, D. G., & Brisson, J. A. (2012). Aphids: a model for polyphenism and epigenetics. *Genetics Research International*, 2012. <https://doi.org/10.1155/2012/431531>

Sparks, J. T., Vinyard, B. T., & Dickens, J. C. (2013). Gustatory receptor expression in the labella and tarsi of *Aedes aegypti*. *Insect biochemistry and molecular biology*, 43(12), 1161-1171. <https://doi.org/10.1016/j.ibmb.2013.10.005>

Srinivasan, D. G., & Brisson, J. A. (2012). Aphids: a model for polyphenism and epigenetics. *Genetics Research International*, 2012. <https://doi.org/10.1155/2012/431531>

Stern, David L. 2008. "Aphids." *Current Biology*: CB 18 (12): R504-5. <https://doi.org/10.1016/j.cub.2008.03.034>

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*, 34(suppl_2), W609-W612. <https://doi.org/10.1093/nar/gkl315>

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology*, 56(4), 564-577. <https://doi.org/10.1080/10635150701472164>

Thissen, D., Steinberg, L. and Wainer, H. (1988). Use of item response theory in the study of group difference in trace lines. In H. Wainer and H. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Thorpe, P., Escudero-Martinez, C. M., Cock, P. J., Eves-van den Akker, S., & Bos, J. I. (2018). Shared transcriptional control and disparate gain and loss of aphid parasitism genes. *Genome biology and evolution*, 10(10), 2716-2733. <https://doi.org/10.1093/gbe/evy183>

Van Dongen, S., & Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. *Bacterial molecular networks: Methods and protocols*, 281-295. https://doi.org/10.1007/978-1-61779-361-5_15

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., ... & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12), 973-982. <https://doi.org/10.1038/nrg2165>

Williams, I. S., & Dixon, A. F. (2007). Life cycles and polymorphism. *Aphids as crop pests*, 69-85.

White, P. S., Arslan, D., Kim, D., Penley, M., & Morran, L. (2021). Host genetic drift and adaptation in the evolution and maintenance of parasite resistance. *Journal of Evolutionary Biology*, 34(5), 845-851. <https://doi.org/10.1111/jeb.13785>

Whiteman, N. K., & Pierce, N. E. (2008). Delicious poison: genetics of *Drosophila* host plant preference. *Trends in Ecology & Evolution*, 23(9), 473-478. <https://doi.org/10.1016/j.tree.2008.05.010>

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, 24(8), 1586-1591. <https://doi.org/10.1093/molbev/msm088>

Zytynska, S. E., Frantz, L., Hurst, B., Johnson, A., Preziosi, R. F., & Rowntree, J. K. (2014). Host-plant genotypic diversity and community genetic interactions mediate aphid spatial distribution. *Ecology and evolution*, 4(2), 121-131. <https://doi.org/10.1002/ece3.916>

4.1 CHAPTER 1 : CHEMOSENSORY GENES AND TRANSPOSABLE ELEMENTS EVOLUTIONARY DYNAMICS IN APHIDS (MANUSCRIPT UNDER SUBMISSION)

Chemosensory gene and transposable element evolutionary dynamics in aphids

Sergio Gabriel Olvera-Vazquez^{1, a)}, Xilong Chen^{1, b)}, Aurélie Mesnil^{1, c)}, Camille Meslin^{2, d)}, Fabricio Almeida-Silva^{3, e)}, Johann Confais^{4, f)}, Yann Bourgeois^{5, g)}, Célia Lougmani^{1, h)}, Karine Alix^{6, i)}, Nathalie Choisne^{4, j)}, Stephane Cauet^{7, k)}, Jean-Christophe Simon^{8, l)}, Christelle Buchard^{8, m)}, Nathalie Rodde^{7, n)}, David Ogereau^{x, o)}, Claire Mottet^{10, p)}, Alexandre Degrave^{11, q)}, Elorri Segura^{10, r)}, Barrès Benoît^{10, s)}, Emmanuelle Jacquin-Joly^{2, t)}, William Marande^{7, u)}, Dominique Lavenier^{12, v)}, Fabrice Legeai^{12, w)}, Amandine Cornille^{1, x)}

¹⁾ Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE –Le Moulon, Gif-sur- Yvette, France

²⁾ INRAE, Sorbonne Université, CNRS, IRD, UPEC, Université Paris Cité, Institut d'Ecologie et des Sciences de l'Environnement de Paris (iEES-Paris), 78026 Versailles, France

³⁾ Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium

⁴⁾ Plant Bioinformatics Facility, BioinfOmics, INRAE, Université Paris-Saclay, 78026 Versailles, France

⁵⁾ DIADE, University of Montpellier, CIRAD, IRD, Montpellier, France

⁶⁾ Génomique Evolutive et Adaptation des plantes Domestiquées-GEvAD, GQÉ-Le Moulon, Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Gif-sur-Yvette, France

⁷⁾ Centre National de Ressources Génomiques Végétales-CNRGV, INRAE, Castanet, Tolosan, France

⁸⁾ Institut de Génétique, Environnement et Protection des Plantes (IGEPP), Domaine de la Motte - 35653 Le Rheu Cedex – France

⁹⁾ Laboratoire Évolution, Génomes, Comportement et Écologie, 1198 Gif-Sur-Yvette Cedex, France

¹⁰⁾ Unité Résistance aux Produits Phytosanitaires, Laboratoire de Lyon, ANSES, Lyon, France

¹¹⁾ Université Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers, France

¹²⁾ IRISA / INRIA. Campus de Beaulieu. 35042 Rennes cedex

- a) sergio.olvera-vazquez@universite-paris-saclay.fr
- b) xilong.chen@inrae.fr
- c) aurelie.mesnil@universite-paris-saclay.fr
- d) camille.meslin@inrae.fr
- e) fabricio.almeidasilva@psb.vib-ugent.be
- f) johann.confais@inrae.fr
- g) yann.bourgeois@ird.fr
- h) celia.lougmani@ universite-paris-saclay.fr
- i) karine.alix@inrae.fr
- j) nathalie.choisne@inrae.fr
- k) stephane.cauet@inrae.fr
- l) jean-christophe.simon@inrae.fr
- m) christelle.buchard@inrae.fr
- n) nathalie.rodde@inrae.fr
- o) david.ogereau@egce.cnrs-gif.fr
- p) claire.mottet@anses.fr
- q) alexandre.degrave@agrocampus-ouest.fr
- r) elorri.segura@anses.fr
- s) benoit.barres@anses.fr
- t) emmanuelle.joly@inrae.fr
- u) william.marande@inrae.fr
- v) lavenier@irisa.fr
- w) fabrice.legeai@inrae.fr
- x) amandine.cornille@cnrs.fr

Corresponding author: amandine.cornille@cnrs.fr

Abstract

Investigating the selection pressures on key genes following duplications and the role of transposable elements (TEs) is key for understanding the evolutionary pace of host-parasite interactions and the associated genomic bases. Chemoreception plays a crucial role in insect adaptation to their plant host. The senses of olfaction and gustation are mediated by large gene families whose evolution is still not fully understood in phytophagous insects. Likewise, the role of TEs in driving the evolution of chemoreception genes still needs to be documented. We characterized chemosensory receptor gene family evolution, TE diversity, and evolutionary dynamics in aphids, which are model systems to understand genomic bases of adaptation of phytophagous insects to their host. We annotated 521 olfactory and 399 gustatory receptor genes (OR and GR) in the genomes of 12 aphid species and one aphid-like species, together with their TE content. We observed a higher evolutionary rate in an aphid subfamily known for its large host breadth and including significant crop pests, driven by the expansion and contraction of genes with functions related to aphid adaptation to their hosts (lipid molecules, immune responses, and transposase activity). Related to the chemosensory genes, we showed low synteny conservation of OR and GR genes between the different species. They evolved by tandem repeats, mainly under relaxed purifying selection compared to other genes in the respective genomes. About 20% of the OR and GR genes evolved under recent positive selection. A negative correlation between rates of evolution of GR/OR genes and species divergence time indicated that recent positive selection is observed in recently diverged sequences. Our results suggest that periods during which OR and GR genes evolve under positive selection may occur episodically and be followed by prolonged episodes of purifying selection. Our results indicated that TEs did not fuel OR nor GR gene evolution. However, we revealed species-specific TE evolutionary dynamics, with major TE bursts in aphid species that infest fruit trees. This study provides insight into the evolutionary processes shaping the evolution of key genes involved in phytophagous insect–plant interactions. It also offers an unprecedented database for studying aphid genome evolution and adaptation to their host.

Keywords: insect, aphid, adaptation, chemoreception, speciation, crop, transposable elements

1 **Introduction**

2 Studying the genomic bases of parasite adaptation to their host can contribute to
3 understanding the evolutionary processes underlying population adaptation (Gladieux et al.,
4 2014; Simon et al., 2015; Möller and Stuckenberg, 2017) and help tackle pressing issues such
5 as parasite outbreaks (Kirk et al., 2013; Cacciò et al., 2018; Desvignes et al., 2022; Gowler et al.,
6 2023). The questions related to parasite adaptation to their host pertain to the identification
7 of the number and nature of genes involved, their mode of evolution by drift (Barton, 1996;
8 White et al., 2021; Lynch, 2023), or natural selection (Bustamante et al., 2005; Fay and Wittkop,
9 2008; Karlsson et al., 2014; Arnab et al., 2023; Culver et al., 2023), and, recently, the role of
10 repetitive regions of the genomes called transposable elements (TEs hereafter; Gilbert et al.,
11 2021) in driving the genomic variation at loci fueling the arms race between parasites and their
12 host.

13 Gene duplication is a source of genomic variation and novel gene functions (Wagner,
14 1998; Magadum et al., 2013). Newly duplicated gene coding can have redundant functions
15 immediately, and one of the copies could accumulate deleterious mutations and become a
16 pseudogene (pseudogenization; Assis and Bachtrog 2013). A duplicated gene can also be fixed
17 by positive selection. When the parasite gene product is functionally divergent, the host
18 cannot recognize it, improving the parasite's fitness. However, periods during which genes
19 evolve under positive selection may occur episodically and be followed by prolonged episodes
20 of purifying selection (leading to an average rate of sequence evolution, $d_N/d_S < 1$ (Yang, 2007;
21 Persi et al., 2016) on a long-term evolutionary time scale. Therefore, investigating the selection
22 pressures on genes following duplications can contribute to understanding the evolutionary
23 pace of host-parasite interactions and the associated genomic bases. In addition to duplicated
24 genes, TEs (mobile repetitive DNA sequences that can independently insert into new genomic
25 regions) can also be involved in parasite adaptation to their host. The selfish proliferation of
26 TEs drives their insertion dynamics, and their impact on host fitness ranges from beneficial to
27 deleterious (Chuong et al., 2017). The fate of TE success in eukaryote genomes depends on
28 past demographic history and selection regimes within species (Lockton et al., 2008; Bourgeois
29 and Boissinot, 2019). Demography can directly impact the efficiency of selection in controlling
30 TE activity and, therefore, genome expansion associated with TE burst (Lynch, 2007). In plant
31 pathogens, positive correlations between genome size and TE content are observed (Lynch,

32 2007). The impact of both demography and selective regimes led to TE-rich genomes
33 displaying genome structure alternating gene-rich and TE-rich regions in plant pathogens. TE-
34 rich genomes have been suggested to facilitate the generation of variation that fuels the arms
35 race between parasites and their hosts (Seidl and Thomma, 2017; Frantzeskakis et al., 2019;
36 Frantzeskakis et al., 2020; Schrader and Schmitz, 2019; Torres et al., 2020). However, diverse
37 types of TEs and higher copies in larger genomes than in smaller genomes are not a universal
38 phenomenon (Elliot et al., 2015), and the correlation between demography, genome size, and
39 TE content remains to be tested in other model systems. Therefore, the numbers, nature, and
40 mode of gene evolution and the associated evolutionary dynamics and role of TEs have been
41 primarily studied in plant pathogens (Duplessi et al., 2011; Plissonneau et al., 2017; Hartmann
42 et al., 2017; Rocafort et al., 2022), but much less characterized in phytophagous insects
43 (Stergiopoulos and de Wit, 2009; Simon et al., 2015; Torres et al., 2020; Seong and Krasileva,
44 2023).

45

46 The olfactory and gustatory receptors (ORs and GRs, respectively) play a major role in
47 host recognition in phytophagous insects (Kambere and Lane, 2007; De Bruyne and Baker,
48 2008; Brand et al., 2015; Rinker et al., 2019; Robertson et al., 2019; Santamaria et al., 2018).
49 These transmembrane proteins are expressed in insect peripheral olfactory and taste neurons.
50 Upon interaction with odorants or tastants, they transform these chemical signals into
51 electrical signals transmitted to the central nervous system. ORs and GRs (Niimura and Nei et
52 al., 2008; Sanchez-Garcia et al., 2009; Ramdya and Benton, 2010; Andersson et al., 2015;
53 Benton, 2015) directly mediate insect interactions with plant cues and chemical defenses and,
54 therefore, are key players in the successful host plant colonization by insects. The ORs are
55 seven-transmembrane domain proteins that participate in the semiochemical conversion into
56 an electrical signal and are expressed in the olfactory neurons located within the olfactory
57 sensilla (Clyne et al., 1999; Scott et al., 2001; Benton et al., 2006; Leal, 2013). The GRs are also
58 a seven-transmembrane domain protein that responds to tastants (*e.g.*, sugars, bitter
59 substances) and is situated within the gustatory sensilla (Chyb, 2004; Sparks et al., 2013; Jiang
60 et al., 2015; Mang et al., 2016). OR/GR genes are reported as highly diversified gene families
61 whose rapid evolution is driven by gene duplication and whose fate depends on selection
62 (positive or purifying) and genetic drift (Hallem et al., 2006; McBride, 2007; Sanchez-Gracia et

63 al., 2009; Joseph and Carlson, 2015). In lepidopterans, multiple OR/GR gene gains and a few
64 gene losses have occurred during their evolution, and it has been hypothesized that the gene
65 family expansions of ORs and GRs have helped the adaptation of lepidopteran species to their
66 host plants (Engsontia et al., 2014). Besides, GR expansion is associated with host breadth in
67 polyphagous Lepidoptera, and GR genes display a cryptic signature of positive selection
68 (Meslin et al., 2022). In dipterans, the variation of OR/GR repertoire sizes among *Drosophila*
69 species is explained by species endemism (Gardiner et al., 2008) and their associated small
70 effective population size and, therefore, genetic drift rather than selection by host specificity
71 (Hanson and Stensmyr, 2011). The link between genome size and the duplication events in
72 chemosensory gene evolution has been tested in different species, e.g., a positive correlation
73 was found in *Drosophila* spp (Gardiner et al., 2008), but negative correlations were reported in
74 lepidopterans (Engsontia et al., 2014). In hemipterans, OR and GR genes show recent and rapid
75 expansion in the pea aphid (Smadja et al., 2009), and the high level of host specialization of
76 this aphid species possibly explains the footprint of positive selection. Studies on the evolution
77 of ORs and GRs in insects have been done on a few model species. This is mainly because
78 studying OR and GR evolution is challenging, requiring manual annotations and time-
79 consuming curations of these OR and GR genes because of their sequence divergence
80 between and within species. Subsequently, the relationship between TEs and chemosensory
81 gene evolution remains poorly studied. TEs are abundant in insect genomes (McCullers and
82 Steiniger, 2017; Gilbert et al., 2021). TE content and type can vary among the genomes of insect
83 species (Gilbert et al., 2021). In several species, high transposon densities are associated with
84 duplicated genes involved in insect adaptation (Gilbert et al., 2021), such as insecticide
85 resistance. For instance, in the ant *Cardiocondyla obscurior* Wheeler, ORs are enriched in TE-
86 dense regions (TE islands), suggesting that TEs play an essential role in OR duplications in ants
87 (McKenzie and Kronauer, 2018). TEs are also close to insecticide resistance genes in
88 *Helicoverpa armigera* Hübner (Klai et al., 2020). Therefore, the relationship between TEs,
89 chemosensory genes, and insect host preference remains to be studied beyond a few insect
90 model species. The explosion of genome assembly makes the comparative studies of OR and
91 GR genes, and TE content and dynamics now reachable and timely in insect species with
92 various host preferences.

93

94 Because of their plant host diversity, Aphids are an ideal system for exploring the
95 genomic bases involved in host adaptation (Peccoud and Simon, 2010; Simon et al., 2015) and
96 the role of TEs (Bouallégué et al., 2017; Ahmad et al., 2021). OR and GR genes are key receptors
97 for aphids, which use chemicals to screen their environmental landscape and detect their host
98 (Robertson et al., 2019). The sequencing of several genomes of aphid species and the
99 annotations of OR and GR genes in two aphid models, *Acyrtosiphon pisum* Harris and *Myzus*
100 *persicae* Sulzer, have provided first hints on the diversity and evolution of chemosensory genes
101 and TE content in aphids (Smadja et al., 2009; International Aphid Genomics Consortium, 2010;
102 Mathers et al., 2017; Thorpe et al., 2018; Li et al., 2019; Fernandez et al., 2020; Julca et al., 2020;
103 Biello et al., 2021; Zhang et al., 2022; He et al., 2023). However, comparative studies of ORs and
104 GRs evolution and their diversification rates among diverse aphid species with various host
105 breaths are lacking. Such comparative studies would help understand if host plant breadth
106 drives aphid speciation (Peccoud, 2010), and what are the associated genomic bases. One
107 challenge to address this question is that the degree of host breadth is complex to assess in
108 aphids. Aphids have various host breaths, from specialists depending on a particular host to
109 complete their life cycle to generalists with several hosts (Dixon, 1977; Williams and Dixon.,
110 2007). This scheme can be even more complex, with aphids, highly polyphagous in the summer
111 on herbaceous hosts, being monophagous on the winter primary host. The modality can also
112 change depending on the geographic region. One way to overcome the difficulty of defining
113 host breadth is to find a group of aphids with a large host breadth compared to another group.
114 In this context, the high species diversity and life cycle of aphids within the subfamily Aphidinae
115 compared to other aphid lineages such as Eriosomatinae (48 genera, most of the species have
116 the elm as a primary host) make them ideal models for correlating the evolution of OR and GR
117 genes with aphid adaptation to their host (Moran, 1992; Blackman and Eastop, 2008).
118 Aphidinae encompasses more than 200 genera; most species have a woody perennial species
119 and one or more herbaceous secondary hosts (Blackman and Eastop, 2008). Aphidinae also
120 includes polyphagous species, among which most agricultural aphid pests are found. A
121 comparative analysis of chemosensory genes (*i.e.*, OR and GR) in Aphidinae and close
122 subfamilies would help to test if increased diversification rates are observed and if it can be
123 associated with the expansion of specific gene families. Likewise, comparing TE content and
124 evolutionary dynamics among aphid species would help us understand the role of TEs in the
125 evolution of chemosensory genes in aphids. So far, TE annotations of available aphid genomes

126 are available, but TE content and evolutionary dynamics comparisons are not possible because
127 of the use of different TE annotators/predictors.

128 In this study, we reconstructed the evolutionary history of OR and GR multigene
129 families in a comprehensive sample of aphid species with various host breadth to gain insight
130 into the selective forces that drive OR and GR evolution. We also characterized TE diversity
131 and evolutionary dynamics in the aphid genomes to understand the role of TEs in the evolution
132 of OR and GR genes in aphids. We relied on one genome of the rosy apple aphid, *Dysaphis*
133 *plantaginea* Passerini, that we newly assembled, and available genomes of 11 aphid species
134 belonging to two main subfamilies (Aphidinae and Eriosomatinae) and one aphid-like species.
135 From the 13 genome assemblies, we inferred single- and multiple-copy orthogroups, manually
136 annotated OR and GR genes, and annotated and classified TEs. We answered the following
137 questions: Are rates of evolution higher in the Aphidinae subfamily? Are they associated with
138 contraction and expansion for specific gene families, particularly the OR or GR gene families?
139 What evolutionary processes shape the molecular evolution of GR and OR genes (by
140 duplication, tandem repeat, and selection regimes)? Are there different TE content and
141 evolutionary dynamics among aphid species, *i.e.*, is there any correlation between aphid TE
142 content, genome size, and demographic history in aphids? Are TEs associated with OR/GR
143 gene evolution (e.g., enrichment in TEs next to OR and GR genes)?

144

145 **Results**

146 **Genetic relationships among aphid**

147 We confirm the relationships among the 12 aphid species (from the Aphidinae and
148 Eriosomatinae subfamilies) and one aphid-like species, *Daktulosphaira vitifoliae* Fitch
149 (Phyloxerinae subfamily) used as the outgroup, based on single-copy genes detected with
150 Orthofinder (Emms and Kelly, 2019) as a first step to analyze the evolutionary dynamics of
151 gene repertoires in aphids. We then characterized the OR and GR gene content of each
152 genome. To that aim, we used a newly assembled *D. plantaginea* genome (version 3, see details
153 in Text S1, Tables S1, and S2) and 12 genome assemblies available online, which had high
154 quality (BUSCO completed genes values > 95.7%, Table S3).

155 From the 13 assemblies, OrthoFinder identified 5,868 orthogroups shared by all
156 species, including 891 single-copy orthogroups (Figure S1 and Table S4). The phylogenetic
157 tree built on the 891 single-copy orthogroups (representing 240,958 conserved amino-acid
158 residues) and time scaled with TimeTree was highly statistically supported (Figure 1). The
159 ultrametric tree showed that the woolly apple aphid, *Eriosoma lanigerum* Hausm, was the most
160 divergent species and formed a specific clade separated from the Aphidinae subfamily. In the
161 Aphidinae subfamily, *Aphis* Linneo and *Rhopalosiphum* Koch species formed a clade separated
162 from another clade, including *Myzus* spp Passerini, *D. plantaginea*, *Diuraphis noxia* Kurdjumov,
163 *Sitobion miscanthi* Takahashi, *Acyrtosiphon pisum* and *Pentalonia nigronervosa* Coquerel.
164 Within this latter clade, *P. nigronervosa* was the most ancestral species; *A. pisum* and *S.*
165 *miscanthus* formed a specific clade separated from *Myzus* spp., *D. plantaginea*, and *D. noxia*.
166 The two *Myzus* formed a monophyletic group.

167

168 **Massive gene gains and losses in the Aphidinae subfamily and their Olfactory Receptors** 169 **(OR) and Gustatory Receptors (GR) gene content**

170 We evaluated the evolutionary history of two major chemosensory genes (OR and GR)
171 that are key for adaptation to their host. Following a homology search and manual curation
172 of OR and GR genes in the 13 genomes (Figure S2), we identified 521 OR and 399 GR genes
173 in *A. gossypii*, *E. lanigerum*, *D. noxia*, *D. plantaginea*, *M. cerasi*, *M. persicae*, *P. nigronervosa*, *R.*
174 *maidis*, *R. padi*, and *S. miscanthi* (Table S3); we used 147 OR and 181 GR genes already
175 annotated in *Aphis glycines* Matsumura, *A. pisum*, and *D. vitifoliae* (no additional set of OR and
176 GR genes were detected for these three species). On average, we saw 44 ± 16.85 GR and $51 \pm$
177 12.78 OR genes per species (Table S3, Figure S3), and the number of OR and GR genes was
178 significantly lower only in *D. noxia* ($P < 0.001$ Model 1, Table S5, Figure S3). There was no
179 significant correlation between genome size and OR or GR gene repertoires (Spearman's
180 correlation $\rho_{GR} = 0.23$, non-significant NS hereafter, $\rho_{OR} = -0.02$, NS).

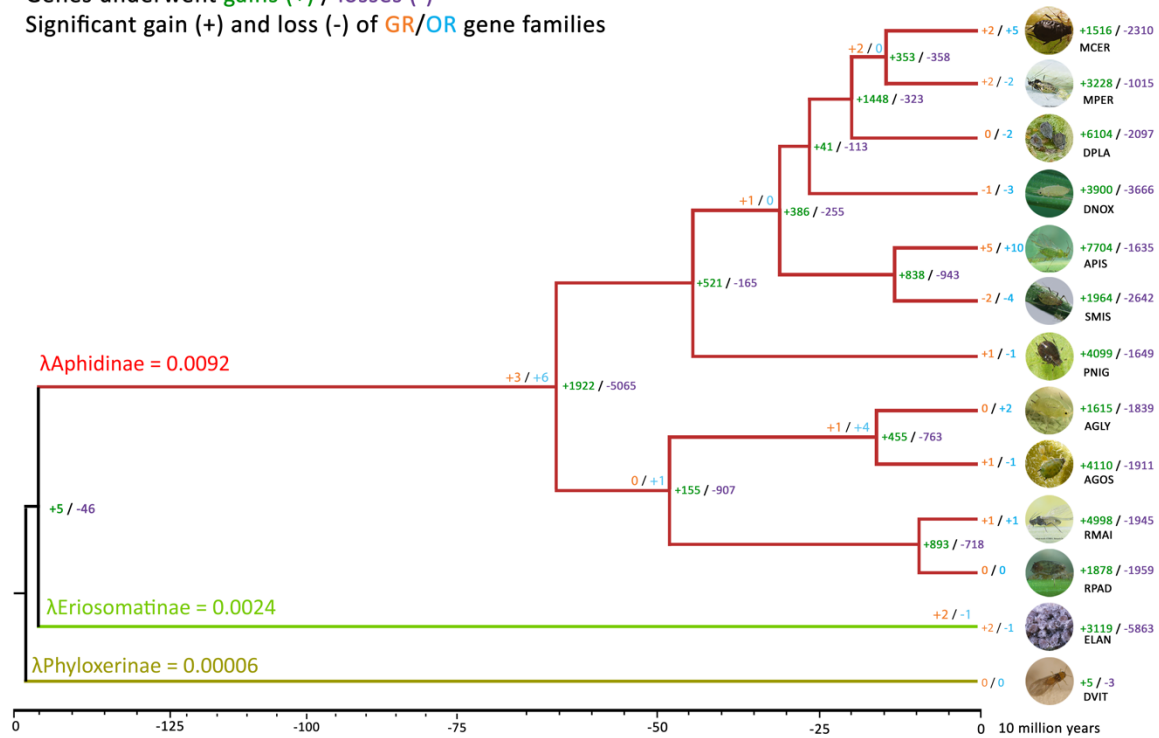
181 Afterward, we used CAFE (Han et al., 2013) to analyze the evolutionary dynamics of
182 gene repertoires in aphids, including OR and GR genes detected above. As inputs for CAFE,
183 we used the ultrametric tree inferred above and the multi-copy orthogroups detected with
184 OrthoFinder. On the 12 aphid species and *D. vitifoliae*, OrthoFinder identified 28,672 multi-

185 copy orthogroups (Figure S1 and Table S4), with 42 orthogroups including 578 GR genes
186 characterized above, 47 orthogroups including 617 OR genes described above, and one
187 orthogroups included both GR and OR genes together with 460 genes. Following the
188 suggestions of the CAFE pipeline, we removed gene copies identified in more than a single
189 species or with less than 100 gene copies to avoid bias in inferences (using a customized
190 Python script from the CAFE pipeline). We, therefore, kept using 22,948 orthogroups, including
191 42 GR and 47 OR orthogroups.

192 We then statistically analyzed the evolutionary rates (gene gain and loss) among gene
193 families (orthogroups) across the phylogeny. We compared models assuming a single
194 evolutionary rate along the phylogeny or evolutionary rates depending on the taxonomic rank
195 (genus, subfamilies, families, species) or host preferences (see details in Table S6). CAFE
196 indicated the highest support for the model with subfamily-specific evolutionary rates ($-lnL =$
197 $243,475$, $\epsilon = 0.02878$, $p\text{-value} < 0.05$; Table S7). For this model, the Aphidinae subfamily
198 showed the highest evolutionary rate (or birth-and-death rate, λ) per copy and million years
199 ($\lambda=0.0092$, $p\text{-value} < 0.05$, Figure 1, Table S7), which was 3.8 times higher than that of the
200 Eriosomatinae subfamily ($\lambda=0.0024$, $p\text{-value} < 0.05$, Figure 1). This result indicates that the
201 Aphidinae subfamily has a higher rate of gain and loss of genes than the Eriosomatinae and
202 Phyloxerinae subfamilies (Figure 1). The internal branch with the most significant number of
203 gene gains (expansions) and loss (contractions) corresponded to the most recent common
204 ancestor of all aphid species from the Aphidinae subfamily (Figure 1). We focused on 147 and
205 51 significantly expanded and contracted gene families ($p\text{-value} < 0.01$) in the Aphidinae
206 subfamily containing at least 50 genes for the gene ontology (GO) analysis. GO analysis of
207 expanded gene families revealed pathways related to metabolism, transcription, lipid
208 molecules and interaction, immune responses, and transposase activity (Figure S5, Table S8).
209 The GO analysis of contracted gene families presented general function (Figure S6, Table S9,
210 $p\text{-value} < 0.05$). Of the 42 GR orthogroups, 11 showed significant expansions and five
211 significant contractions along the 12 aphid species. Six of the 47 OR orthogroups showed
212 significant expansions and nine significant contractions. Specifically, we observed four
213 expanded chemosensory receptors (two GRs and two ORs) in the Aphidinae (Figure 1). We
214 observed the highest gain of GR and OR genes in *A. pisum* and *Myzus cerasii* Fabricius. Our
215 results indicated fewer OR and GR duplication or contraction events than for other gene

216 families described above.

Genes underwent gains (+) / losses (-)
Significant gain (+) and loss (-) of GR/OR gene families



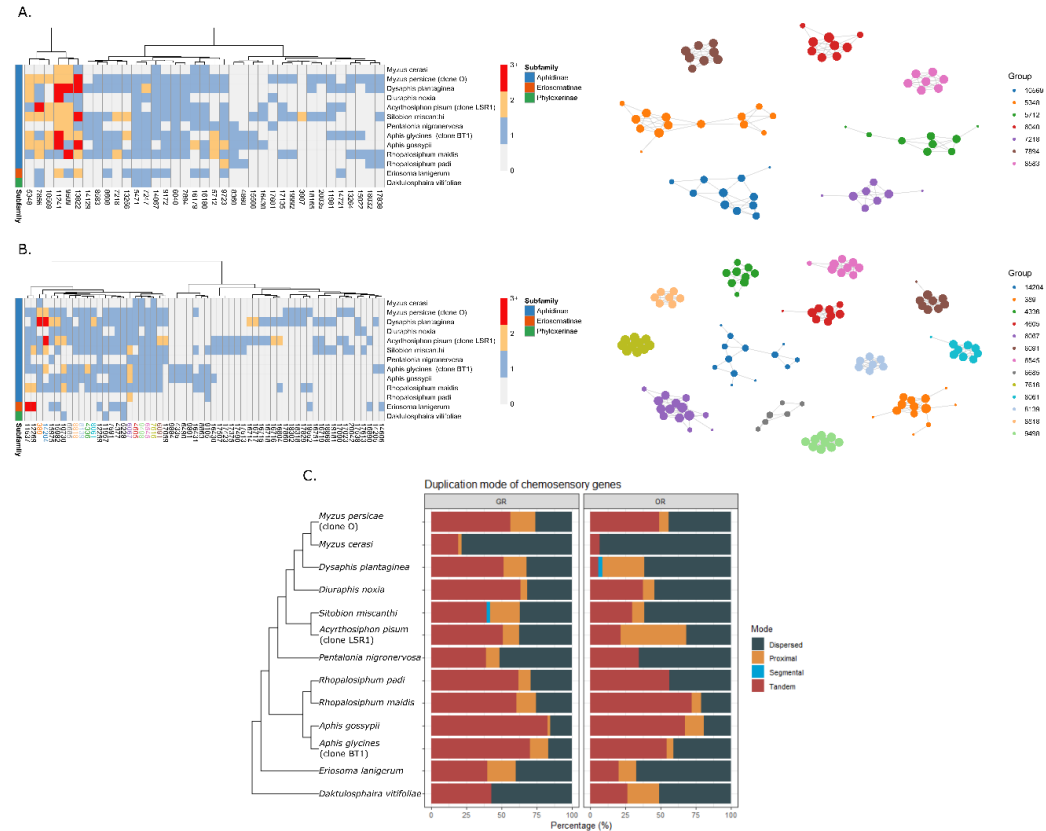
219 **Figure 1.** Gene expansion and contraction events in the 12 aphid species and the outgroup
 220 *Daktulosphaira vitifoliae* detected with CAFE software. AGLY = *Aphids glycines* clone BT1;
 221 AGOS = *Aphis gossypii*; APIS = *Acyrtosiphon pisum* clone LSR1; DNOX = *Diuraphis noxia*;
 222 DPLA = *Dysaphis plantaginea*; DVIT = *Daktulosphaira vitifoliae*; ELAN = *Eriosoma lanigerum*;
 223 MCER = *Myzus cerasi*; MPER = *Myzus persicae* clone O; PNIG = *Pentalonia nigronervosa*; RMAI
 224 = *Rhopalosiphum maidis*; RPAD = *Rhopalosiphum padi*; SMIS = *Sitobion miscanthi*. The figure
 225 represents the number of genes lost or gained during the analyzed species' evolution. All
 226 branches are supported with 100% bootstrap values.

227 **Molecular evolution of GR and OR genes: synteny, mode of evolution and selective**
228 **pressures**

229 We investigated the molecular evolution of OR and GR genes using three approaches.
230 First, we investigated their synteny, then their mode of evolution, and third, the selective
231 pressures underlying their evolution.

232 We investigated microsynteny of OR and GR protein sequences across the 12 aphid
233 genomes with the syntenet R package (Almeida-Silva et al., 2023). This meant preserved order
234 of genes on the same chromosome (syntenic cluster; Duran et al., 2009) among the aphid
235 genomes. Syntenet indicated that 216 out of 668 ORs (33.8%) and 228 out of 580 GRs (39.5%)
236 belonged to syntenic clusters (37 and 60 clusters, respectively) across the 12 aphid species.
237 The phylogenomic profiles revealed that synteny clusters were not deeply conserved across
238 taxa, but the GR microsynteny was more conserved than the OR one within the Aphidinae
239 subfamilies (Figure 2). The syntenic clusters did not group specific species (Table S10). Note
240 that syntenet detected 101 genes with putative functions as chemosensory genes (Table S10).
241 Some of these detected genes were also clustered in orthogroups by OrthoFinder. We found
242 70 and 31 genes related to putative olfactory and gustatory functions, respectively (Table S10).
243 However, we did not consider them for further analyses as these genes were not detected
244 during the manual annotation of OR and GR genes.

245 To investigate the mode of evolution of OR and GR genes, we identified and classified
246 them as dispersed duplicates (generate gene copies neither neighboring nor colinear),
247 proximal duplicates (gene copies separated by 10 or fewer genes on the same chromosome),
248 tandem duplicates (gene copies closely adjacent to each other in the same chromosome), and
249 segmental duplicates (gene copies block with high sequence identity; Qiao et al., 2019;
250 Abdullaev et al., 2021), depending on their copy number and genomic distribution. Most OR
251 and GR genes originated from tandem and proximal duplications (Figure 2). For some species,
252 most genes originated from "dispersed duplication", meaning they could not be classified into
253 any segmental, tandem, or proximal.



254

255 **Figure 2. Phylogenomic profile, syntenic network, and duplication evolution mode of OR and GR encoding genes in aphid.** A) Gustatory
 256 receptors encoding genes. B) Olfactory receptors encoding genes. The heatmaps represent the number of genes in each genome syntenic block
 257 (clusters). Species are sorted according to the aphid subfamily; the species belong to syntenic blocks (clusters) and are grouped with Ward's
 258 clustering on a matrix of Euclidean distances according to the phylogenomic profile of each cluster. Each gene in a syntenic block is represented
 259 as a node, and syntenic relationships between nodes are represented as edges. The less conserved GR cluster (7,305) and OR cluster (13,314) are
 260 found in 54% of Aphidinae, while the more conserved cluster (9,291 for GR and 7,461 for OR) are found in Aphidinae. C) Percentage of
 261 duplication modes of each aphid species.

262 We investigated the signatures of selection in OR and GR genes. To that aim, we
263 classified the subfamilies of OR and GR genes using a Maximum Likelihood tree on 373 OR
264 amino-acid sequences and 393 GR amino-acid sequences from 12 aphids and one aphid-like
265 genome. The GR clustered into 12 subfamilies (monophyletic groups with high bootstrap
266 values > 80%, Figure S7), and the OR clustered into 24 subfamilies (high bootstrap values >
267 80%, Figure S8). We detected two known GR subfamilies, the sugar and fructose receptors
268 (GR01 and GR02, respectively, Figure S7), based on the similarities with the already reported
269 ones in *A. pisum* and *A. glycines*. We also detected the conserved odorant co-receptor (Orco,
270 OR01 in Figure S8) subfamily based on the clustering with the reported Orco genes of *A. pisum*,
271 *A. glycines*, and *D. vitifoliae*. The high number of OR subfamilies compared to GR subfamilies
272 and the star-like shape of the phylogenetic tree (Figure S8) indicated higher OR evolutionary
273 rates than GR genes.

274 We then used the codeml program in the PAML package v4 (Yang, 2007) to determine
275 whether the substitution rate of the protein sequences (ω) of the 12 and 24 GR and OR
276 subfamilies, respectively, evolved neutrally or by natural selection (positive or purifying).
277 Codeml computes the ratio (ω) of nonsynonymous (dN) on synonymous (dS) substitutions for
278 various models of sequence evolution (tree branch or/and codon-site). The most likely model
279 explaining the data is chosen based on a likelihood ratio test computed among models. Here,
280 we used the branch and site models. We did not use the branch-site model as we neither had
281 any previous information about the phylogeny of these genes nor an apparent chemosensory
282 gene family clustering to test for selection. We then retrieved the ratio (ω) of nonsynonymous
283 (dN) and synonymous (dS) estimates from the most likely models and compared them between
284 GR and OR genes, as well as to the rest of the genomes (OG genes; see below).

285 For most GR and OR subfamilies, the free ratio (different ω values, (Yang, 2007)) among
286 branches was the most statistically supported model (Table S11), indicating that within each
287 OR and GR subfamily, sequences evolved with their own ω (Table S12). One of the exceptions
288 was the fructose receptor genes (GR02), which displayed a single evolutionary rate among
289 the sequences (Table S12). Of the 393 GR sequences analyzed, 57.7% evolved under purifying
290 selection, 21.62% under positive selection, and 20.61% under neutral selection. Of the 373 OR
291 sequences, 46.11% evolved under purifying selection, 25.47% under positive selection, and

292 27.34% under neutral evolution. The signatures of selection were not subfamily specific: each
293 subfamily presented each of the three signatures of selection (Tables S11 and S12). However,
294 lineage-specific models (branch models) have limited ability to detect short episodes of
295 positive selection that affect only a few amino acids or amino acids under recurrent diversifying
296 selection. We therefore used site-specific models implemented in codeml on sequences of
297 each GR and OR subfamily. The site-specific models can account for site rate variation and
298 detect recent episodes of positive selection (diversifying selection). We used three pairs of site-
299 specific models to test for recurrent, diversifying selection: M0 (one ratio) and M3 (Discrete),
300 M1 (Neutral) and M2 (Selection), and M7 (Beta) and M8 (Beta & ω). The likelihood ratio test
301 showed higher support for neutral or evolution by purifying selection (M0, M1a, and M7)
302 models than the alternative ones assuming diversifying selection (M3, M2a, and M8) (Tables
303 S13 and S14), except the gene groups GR02 (fructose receptor) and OR03 (Table S14). Further
304 examinations of the posterior probability (PPs > 95%) using the Bayes empirical Bases (BEB)
305 distribution of the inferred positively selected sites under the M8 model in these two group
306 genes suggested 19 sites in the GR02 (seven overlapping sites between M2a and M8) and one
307 site in the OR03 (this site overlapped in M2a and M8; Table S15). These proposed sites should
308 be considered cautiously because the likelihood method used to infer them is known to falsely
309 detect positively selected sites (Suzuki and Nei, 2002; Zhang, 2004). Altogether, branch and
310 site models indicate that OR and GR genes mainly evolved under purifying selection.

311 We compared ω values estimated with the branch models between OR and GR genes
312 and the rest of the genome. We represented the rest of the genome with 50 randomly chosen
313 single-copy genes detected with Orthofinder. A total of 14 % of the single-copy orthologous
314 gene groups evolved under the same ω , while 86 % fitted the branch models with a specific
315 ω (Tables S15 and S16). For the site models, sequences of 49 single-copy orthologous families
316 evolved under the same ω (M0), including 72% evolving under purifying selection ($\omega < 1$; Table
317 S17). We then compared the dN/dS ratio of the single-copy genes to the OR and GR gene
318 subfamilies. The OR and GR genes showed significantly weaker purifying selection than single-
319 copy genes (fittest model, Model 2, Figure S9, Table S20). Our results indicate that GR and OR
320 genes mainly evolve under purifying selection (Figure S9), with nonsignificant differences in
321 selection pressure on OR genes than on GRs. We also observed a negative correlation between
322 the dN/dS ratio and divergence among species, indicating that genes with the lower

323 divergence were under the highest evolutionary rates ($P < 0.0004$, $r^2 = -0.09$, Figure S10).

324 Our results indicate that OR and GR genes evolved mainly under purifying selection
325 but much weaker than the median strength of purifying selection affecting genes in the
326 respective genomes. About 20% of the OR and GR genes evolved under recent positive
327 selection. A negative correlation between rates of evolution of GR/OR genes and species
328 divergence time indicated that recent positive selection is observed in recently diverged
329 sequences.

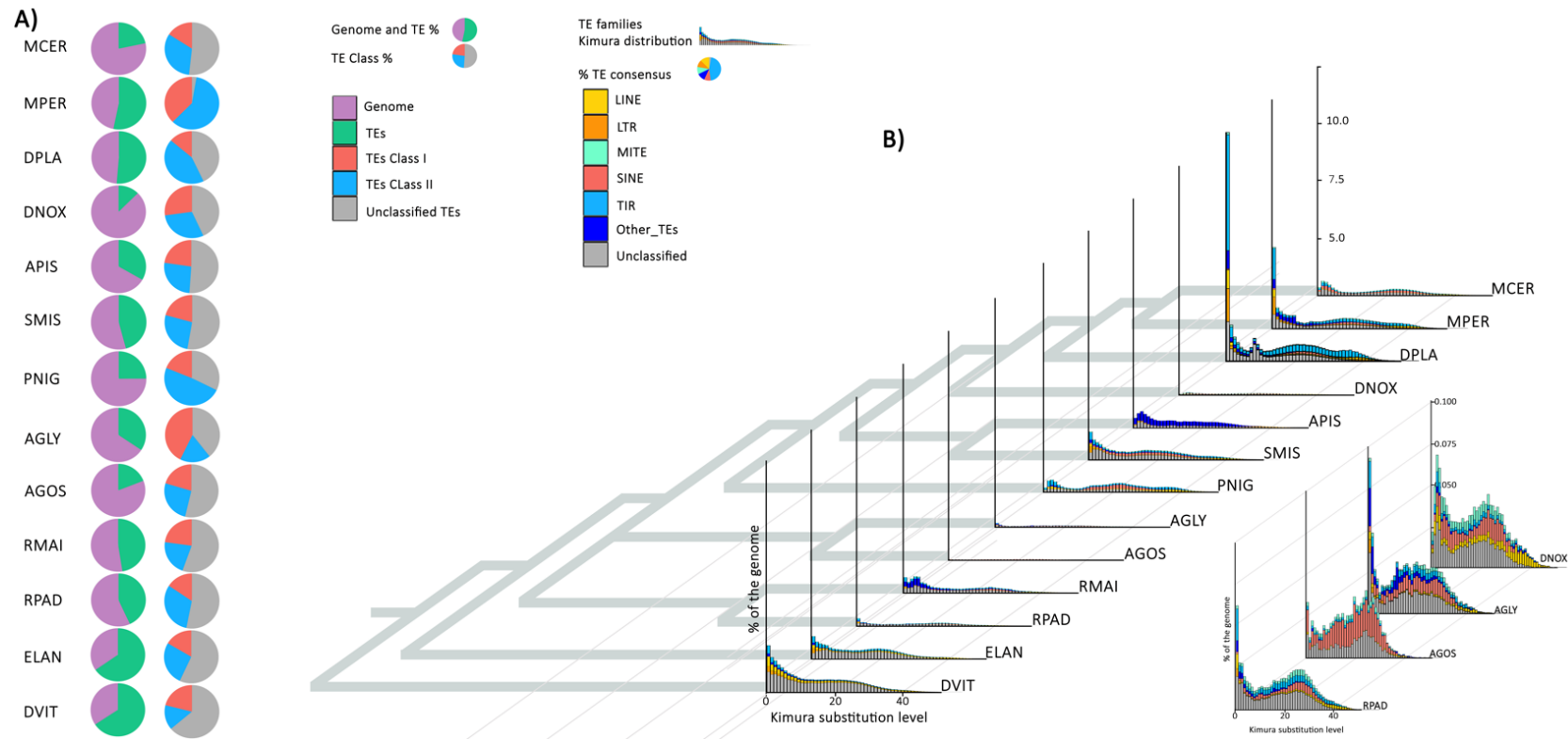
330 **Various TE content and evolutionary dynamics among aphid species**

331 TE prediction and annotation were not significantly correlated with assembly quality
332 (Table S21 and S22). We did not find a correlation between TE content and genome size ($r = 0.02$,
333 NS, Table S6). We observed variation in the proportion of TE types among species, but it was not
334 significant (Model 4, P -value > 0.005 ; Figure 3A). We observed nearly all known TE superfamilies
335 in at least one aphid or aphid-like species with type II and I TEs. TIR TEs were the most abundant,
336 followed by LINE and LTR retrotransposons (Figure 3B). Note that a large "unclassified" portion
337 may mask other TE types. We also observed other TEs in low proportion, such as Crypton,
338 Helitron, and Maverick (Class II), and DIRS, LARD, PLE, and TRIM (Class I).

339 We detected various Kimura distance-based copy divergences of each TE order, indicating
340 different TE evolutionary dynamics among aphid species (Figure 3B). The relative age of a TE
341 family can be seen as its distribution in the graph: the "older" TE families have a more extensive,
342 flatter distribution spreading to the right (higher levels of substitution) than the recent TE families,
343 which are on the left of the graph and have a restricted distribution (Parisot et al. 2021). The
344 Kimura distance represents the calculation between each annotated TE copy and the consensus
345 sequence of the respective TE family. High Kimura distance suggests a TE copy cluster with high
346 sequence divergence due to genetic divergence among TE copies. We found a contrasting peak
347 of low distance in DNA transposons in *D. plantaginea* followed by *M. persicae* (Figure 3B), which
348 indicates a recent TE activity and expansion in these species. The recent bursts of TE in the rosy
349 apple aphid and *M. persicae* mainly concerned TIR. In contrast, four aphid species presented low
350 values of TE divergence, indicating ancient TE dynamics. For instance, *A. gossypii*, *D. noxia*, *A.*
351 *glycines*, and *R. padi* displayed values from 0.025 to 0.100, reflecting ancient historical TE activity
352 in the genomes of these species (Figure 3). To analyze the relation between historical population

353 dynamics and the transposable elements we tracked the historical fluctuations in species effective
354 population size (N_e) inferred with the pairwise sequentially Markovian coalescent approach
355 (PSMC; Li and Durbin, 2011) using masked and unmasked genomes (Figure S11). The estimates
356 of N_e for the masked genomes compared to the unmasked genomes were substantially different
357 for *D. plantaginea*, *D. vitifolia*, *E. lanigerum*, and *M. persicae*.

358 The intersection between TE copy cluster content among the different aphid species
359 revealed a phylogenetic signal in shared TE copy (Figure S12). The highest number of TE copies
360 was shared between *M. persicae* and *D. plantaginea*, with TIR being the most abundant, followed
361 by TRIM, LTR, Line, and Helitron. The aphids *R. maidis* and *R. padi* shared a higher TE content of
362 TRIM, followed by TIR, MITE, LINE, Helitron, and PLE (Figure S12).



374

375 **Figure 3.** Transposable element (TE) content among twelve aphid species and the grape phylloxera. A) Proportion of TEs in the genome and
 376 consensus TEs of Class I and II of each analyzed species. B) TE kimura distribution among the analyzed species. AGLY = *Aphids glycines* clone BT1;
 377 AGOS = *Aphis gossypii*; APIS = *Acyrtosiphon pisum* clone LSR1; DNOX = *Diuraphis noxia*; DPLA = *Dysaphis plantaginea*; DVIT = *Daktulosphaira*
 378 *vitifoliae*; ELAN = *Eriosoma lanigerum*; MCER = *Myzus cerasi*; MPER = *Myzus persicae* clone O; PNIG = *Pentalonia nigronervosa*; RMAI =
 379 *Rhopalosiphum maidis*; RPAD = *Rhopalosiphum padi*; SMIS = *Sitobion miscanthi*.

380 **A few TE enrichments but no depletion events near OR/GR genes**

381 We investigated TE enrichment and depletion near OR and GR genes using the LOLA
382 R package (Sheffield and Bock, 2016) to assess the role of TEs in the evolution of the GR and
383 OR genes in aphids. We found significant TE enrichment (q -values < 0.05) in the surrounding
384 of chemosensory receptor genes in only 3.30% of OR and 0.5% of GR genes. Some notable
385 examples of TE enrichments are found in four species when extending the universe (genomic)
386 and query regions (OR and GR) to 10Kb: *A. glycine*, *D. plantaginea*, *D. vitifoliae*, and *R. padi*
387 (Table S22). In *A. glycines*, we found unclassified TE enrichment near three GR and two OR
388 genes. In *D. plantaginea*, we observed TIR enrichment near three OR genes. *Rhopalosiphum*
389 *padi* presented more TIR, LINE, and unclassified enrichment in twelve OR genes. Finally, we
390 found unclassified TE enrichment in OR genes in *D. vitifoliae*. We found a lack of TE depletion
391 near the OR/GR genes (Table S23). We also evaluated TEs that would be related to gene
392 regulation, limiting the universe (genomic) and query regions (OR and GR) to 2 Kb. In insect
393 gene studies, promoters were detected at 2Kb from the start codon (Liu and Zhang, 2002;
394 Simola et al., 2013; Chen et al., 2020; Liu et al., 2021). We detected TE enrichment of SINE/LINE,
395 unclassified, and TIR related to ORs in *A. pisum*. We also detected TE enrichment of unclassified
396 and TIR linked to ORs in *D. plantaginea*. We also found TEs related to unclassified close to ORs
397 in *D. vitifoliae*. In *S. miscanthi*, we found unclassified TEs associated with GRs and ORs. We also
398 found unclassified TE enrichment in the ORs of *R. padi*. On the other hand, we found support
399 for TIR TE depletion in *R. maidis* related to GRs (Table S24). Altogether, we found a few TE
400 enrichments near the GR and OR genes.

401

402 **Discussion**

403 We utilized comparative genomics to examine the evolution of genes putatively
404 involved in host-phytophagous insect interactions. We manually annotated olfactory and
405 gustatory receptor genes in 12 aphid species. These genes evolved in tandem repeats, mainly
406 under purifying selection at that long-term evolutionary timescale. We showed species-
407 specific TE content and evolutionary dynamics with a recent dynamic in species-infesting fruit
408 trees (*D. plantaginea* and *M. persicae*). We showed that TEs did not cluster near GR and OR

409 genes, suggesting that TEs did not fuel the diversity of ORs and GRs at that evolutionary time
410 scale. This study provides insight into the evolutionary processes shaping the dynamic of key
411 genomic elements of major crop pest genomes to adapt to their host. It also offers an
412 unprecedented database for studying aphid genome evolution and aphid adaptation to their
413 host.

414

415 **The Aphidinae subfamily shows a high evolutionary rate with an expansion of key genes**
416 **involved in interactions with their plant hosts.**

417 Previous studies have hypothesized a relationship between the host range and
418 speciation rates (Nicholson et al., 2015; Robertson, 2019). Assessing the degree of host breadth
419 or specialization in aphids is complex as it depends on the time of the year and the geographic
420 region. To overcome this limit, we studied the evolutionary history of aphid species from the
421 Aphidinae subfamily, which includes many highly polyphagous species with various host
422 breadths, to study whether host plant breadth drove aphid speciation in this subfamily
423 (Peccoud, 2010). We considered aphid species reported as polyphagous or generalists with a
424 high host range to try to study gene evolution in a highly diversified family. For instance, *A.*
425 *gossypii* feeds on about 900 plants (Blackman and Eastop, 2000; Ma et al., 2019), *M. persicae*
426 on around 400 plant species, and *D. noxia* on at least 140 plant species (Blackman and Eastop,
427 2000). We also included polyphagous species with more host restriction, like the model species
428 *A. pisum* feeding on plants from the Fabaceae family and *A. glycines* feeding on *Glycine* Willd.
429 spp and *Rhamnus* L. spp (Ragsdale et al., 2004; Wang et al., 2019). We added aphid species
430 that infest mainly the Poaceae family, including essential cereal crops such as maize, barley,
431 oat, and wheat; these species were *R. maidis* (Chen et al., 2019), *R. padi* (Thorpe et al., 2018),
432 and *S. miscanthi* (this species also infests Cyperaceae members; Jiang et al., 2019). On the other
433 hand, we added more specialized aphids, *D. plantaginea* and *E. lanigerum*, whose primary host
434 is the apple tree (*M. domestica* Borkh); note that *E. lanigerum* has the elm as the main host in
435 the USA. *Myzus cerasi* (Dondini et al., 2018) infests a limited host range, including cherry trees
436 (*Prunus avium* L. and *Prunus cerasus* L.). The aphid *P. nigronervosa* infests *Musa* L. spp (Mathers
437 et al., 2020), and we also included the aphid-like species *D. vitifoliae*, which feeds on different
438 species of the *Vitis* L. genus (Rispe et al., 2020).

439 First, we obtained a phylogenetic tree that fits the taxonomy of the species like other
440 publications (Choi et al., 2018; Mather et al., 2020; Byrne et al., 2022; Huang et al., 2023). We
441 confirmed the monophyletic group of the tribe Macrosiphini (*M. cerasi*, *M. persicae*, *D.*
442 *plantaginea*, *D. noxia*, *A. pisum*, *S. miscanthi*, and *P. nigronervosa*), Aphidini (*A. glycine*, *A.*
443 *gossypii*, *Rhopalosiphum maidis* Fitch, and *R. padi*), Eriosomatini (*E. lanigerum*), and Phylloxerini
444 (*D. vitifoliae*). Second, we showed higher evolutionary rates driven by expansions and
445 contractions of gene families related to the interaction of aphids with their hosts (lipid
446 molecules and interaction, immune responses, and transposase activity) in the Aphidinae
447 subfamily known for its large host breadth. Clades encompassing species that exhibit large
448 host breadths in insects are expected to have higher speciation rates (Hardy and Otto 2014;
449 Weingartner et al. 2006). Our results, therefore, validate this hypothesis, albeit additional
450 genomes outside the Aphidinae family would reinforce our results. Besides, pseudogenisation
451 events could be further investigated across the phylogeny. These genomic fossils or
452 evolutionary relicts can help better understand each genome evolution rate, as they generally
453 evolve without selective constraints (Podlaha and Zhang, 2010).

454

455 **The molecular evolution of the olfactory and gustatory receptors followed the birth-**
456 **and-death model of evolution and evolved mainly under purifying selection.**

457 Chemosensory receptor genes play critical roles in detecting chemical cues in insects
458 (Benton et al., 2006; Smadja et al., 2009) and driving behaviors like mating, oviposition, and
459 feeding (Li et al., 2017). We, therefore, analyzed the evolutionary history and molecular
460 evolution of olfactory and gustatory receptors since they are involved in the host-aphid
461 interaction. Despite their variable host preferences, we did not find variation in the OR/GR
462 gene repertoire size among aphid species, but we found various evolutionary dynamics. We
463 also revealed the mode of evolution of OR and GR genes.

464 As we used a standardized pipeline on high-quality reference genomes, this lack of
465 differences may not be due to the missing OR and GR genes. Quan et al. (2019) showed a
466 lower number of GR genes in generalist species (*A. gossypii* 50 GRs, *M. persicae* 56 GRs, and *D.*
467 *noxia* 30 GRs) than in specialist aphid species (*A. glycines* 64 GRs and *A. pisum* 77 GRs), but

468 there was no difference in OR repertoire size (*A. gossypii* 34 ORs, *M. persicae* 50 ORs, *D. noxia*
469 31 ORs, *A. glycines* 37 ORs, and *A. pisum* 79 ORs). Robertson et al. (2019) classified *A. pisum* as
470 a generalist and *A. glycines* as a specialist and showed a higher number of OR (87) and GR (78)
471 genes in *A. pisum* than in *A. glycines* (47 ORs and 61 GRs). They conclude that the numbers of
472 OR and GR genes are due to the host breadth of each aphid species. Various conclusions were
473 drawn in other insect species with different host specializations, such as the specialist fig wasp
474 *Ceratosolen solmsi* Mayr which displays 46 OR and 6 GR genes (Xiao et al., 2013), while the
475 specialized human body louse *Pediculus humanus* Haeckel has 10 ORs and 6 GRs (Kirkness et
476 al., 2010). Here, we used more aphid species than in previous studies and a standardized
477 annotation pipeline of OR and GR genes. We also stressed in the introduction the difficulty of
478 defining specialist and generalist species in aphids and, therefore, it is complicated to draw
479 conclusions related to host specialization in aphids as it relates to a vague definition of
480 specialization.

481 However, our results indicated various evolutionary dynamics of OR and GR genes. The
482 GR gene family is assumed to be the most ancient chemoreceptor family that has evolved in
483 insects (Robertson et al., 2003; Vieira and Rozas, 2011). The OR gene families evolved from an
484 ancestral GR family and are reported to be a highly diverged group (Brand et al., 2018). For
485 instance, highly divergent clades were detected by building a phylogenetic tree of 62 ORs and
486 68 GRs in *D. melanogaster* (Robertson et al., 2003). We confirmed this pattern in the OR start-
487 like phylogenetic tree. Our results also show species-specific OR and GR genes but not massive
488 GR and OR gain and losses, except for *A. pisum*. Our results suggest that GR and OR relay to
489 ancestral function were maintained. Various expansions of OR and GR genes were observed in
490 insects, sometimes associated with the adaptation to multiple hosts such as *Apolygus lucorum*
491 or *Drosophila suzukii* (Hickner et al., 2016; Liu et al., 2021). The significant chemosensory gene
492 expansion in *D. suzukii* is explained by the range of life history traits, from feeding on primitive
493 sap and slime to decaying and fermented fruits (Hickner et al., 2016). We found more
494 expansion and contractions of OR and GR in *A. pisum*, probably reflecting the large host
495 preferences, but analysis using more host representatives should be added. Our results
496 showed that in the Aphidinae family, which includes many species with large host breadth,
497 major expansions of gene families related to interactions of aphids with their hosts (lipid

498 molecules, immune responses, and transposase activity) are observed. Several gene families
499 could, therefore, be studied in the future to further dig into the genomic bases of adaptation
500 in aphids. This would require a standard gene annotation pipeline for the concerned gene
501 families (e.g., effector families such as salivary effectors or transposon family).

502 We provided insights into the molecular evolution of OR and GR genes. We revealed
503 that OR and GR genes show low micro-synteny, evolved by tandem and proximal duplications,
504 and mainly under purifying selection. This duplication mode has suggested rapid divergence
505 and strong selective pressures (Qiao et al., 2019). This tandem duplication evolution was
506 observed in *Drosophila yakuba* Burla, and *Drosophila simulans* Sturtevant related to genes
507 involved in defense against pathogens, insecticide resistance, chorion development, cuticular
508 peptides, and lipases (Rogers et al., 2015). However, the mechanism is unclear; in *Arabidopsis*
509 *thaliana* L., it is proposed to facilitate gene coexpression via "gene block" improved RNA
510 polymerase transcription due to the physical proximity (Amoutzias and Van de Peer, 2008). We
511 also observed that the ancestral functions of the chemosensory genes have been maintained
512 by the retention of the sequences and by excluding deleterious mutations by purifying
513 selection (Diepeveen et al., 2013). A part of the OR and GR sequences in aphids, however,
514 evolved by positive selection, and we observed a negative correlation between dS , suggesting
515 that genes that evolve under positive selection may occur episodically and be followed by
516 prolonged episodes of purifying selection (leading to an average $dN/dS < 1$). The timescale of
517 our studies among the aphid genus may explain the lack of detection of recent positive
518 selection, as observed, for instance, in *A. pisum* at the within-species level (Smadja et al., 2009).
519 A population-level study of the selection pattern in candidate genes or through genome scans
520 is needed further to understand the recent evolutionary dynamics of aphid genomes.

521

522 **Few enrichments of TEs in OR and GR genes, but recent TE activities on aphids on fruit** 523 **trees**

524 TE dynamics did not correlate with genome size, suggesting that not only demography
525 and purifying selection are shaping the TE landscape in aphids. As discussed in other studies,
526 several possible explanations for these complex patterns may require further investigations,

527 including horizontal gene transfer (Elliot and Gregory, 2015).

528 Transposons may have fueled the diversity of one or a few OR/GR genes, but they are
529 most likely not the main source of chemosensory genes at the evolutionary scale of our study.
530 We tested the TE enrichment and depletion near OR and GR genes. We found very few, and
531 mostly in OR genes, enrichment of TEs. The evolutionary timescale of our study can explain
532 our lack of support for TE enrichment near OR and GR genes. TE could have played a role in
533 fueling the diversity of ORs and GRs, but they may have been eliminated by purifying selection
534 once their potential in driving OR/GR diversity was effective. This may explain the few signals
535 observed, mainly next to OR genes, which are known to be a more recent gene family than GR
536 genes, as indicated by the star-like phylogeny of the OR genes. It has recently been shown
537 that the proportion of unclassified TEs is high for *A. pisum* used in this study (Sproul et al.,
538 2023). Almost 50% of the TEs analyzed were unclassified in our research, which is close to what
539 is observed for non-model species. However, we may have missed TE families, particularly
540 short-length TEs.

541 We observed interspecies variation in TE evolutionary dynamics in aphids. The relative ages
542 of TE were heterogeneous among species, suggesting contrasting TE activity among species, with
543 very recent activity in species found on fruit trees, *D. plantaginea*, and *M. persicae*. The two species
544 are found on fruit trees and have recent evolutionary histories (Olvera-Vazquez et al., 2021; Singh
545 et al., 2021). However, it is too early to link TE evolutionary dynamics and the species' ecology or
546 evolutionary history. Further studies at the population level are required. The TE heterogeneous
547 distribution along genomes may lead to SNPs masking in regions of low and high genetic
548 diversity, altering the PSMC. We made our PSMC inferences using a single individual per species;
549 the population genomics lens of the transposon insertion polymorphism within species,
550 particularly in the species showing recent activity, such as *D. plantaginea*, would help to see if
551 transposon insertion is observed near OR and GR genes. Therefore, more investigations of the
552 role of TEs in the evolution of each species using a population genomics lens would provide
553 insights into the role of TEs in aphid evolution and, more generally, on the balance between
554 demography and selection in insect evolution (Lockton et al., 2008; Oggenfus et al., 2021).

555

556 **Conclusion**

557 Our comparative genomics analysis suggests that different evolutionary forces on OR
558 and GR genes may dominate different evolutionary scales. Purifying selection may dominate
559 on the long timescale, while positive selection may dominate on a short time scale.
560 Demography may influence recent TE evolution, while other processes may more strongly
561 determine phylogenetic-scale patterns in TE landscapes. This study, therefore, provides the
562 first insights into the evolutionary dynamics and molecular evolution of key genes possibly
563 involved in the adaptation of phytophagous insects to their host. We also provide a high-
564 quality genome assembly of the rosy apple aphid (*D. plantaginea*) and an unprecedented
565 database of annotated ORs, GRs, and TEs, as valuable supports for studying aphid genome
566 evolution and aphid adaptation to their host. Now, further studies are needed at the
567 population level.

568

569 **Material and methods**

570 **Analyzed genomes**

571 We assembled a high-quality reference genome of *D. plantaginea*, the rosy apple aphid
572 (see details in Text S1, Table S1, and S2). We retrieved the genomes of twelve aphid species
573 and one aphid-like species from previous publications (Table S3): *Aphis glycines* (bt1),
574 *Acyrtosiphon pisum* LSR1 (v2), *Aphis gossypii*, *Diuraphis noxia*, *Eriosoma lanigerum*, *M. persicae*
575 (clone O), *Myzus cerasi*, *Pentalonia nigronervosa*, *Rhopalosiphum maidis*, *Rhopalosiphum padi*,
576 and *Sitobion miscanthi*, and one aphid-like species of the Phylloxeridae family, *Daktulosphaira*
577 *vitifoliae* Fitch. The quality of the genomes was assessed with the *N50* (shortest contig length
578 that needs to be included for covering 50% of the genome) and BUSCO (Benchmarking
579 Universal Single-Copy Orthologs) scores detailed in the publications (Table S3). We
580 downloaded the genome assemblies and annotations of the aphid species from the aphid
581 genome database AphidBase – Bioinformatics platform for Agroecosystem Arthropods (BIPAA;
582 Legeai et al., 2010: <https://bipaa.genouest.org/is/aphidbase/>).

583

584 **Annotation of chemosensory genes**

585 We used the previously reported annotated chemosensory genes (ORs and GRs) for *A.*
586 *glycines*, *A. pisum*, and *D. vitifoliae* (Robertson et al., 2019; Rispe et al., 2020; Table S3) as input
587 for Exonerate (Slater and Birney, 2005), the InsectOR pipeline (Karpe et al., 2021), and Scipio
588 (Keller et al., 2008) for gene prediction of both chemosensory gene families in the 13 genomes
589 described above (Figure S2). We manually curated and filtered the gene models predicted on
590 the Apollo server (Lee et al., 2013) installed on the BIPAA platform, blasting on NCBI to identify
591 similarity percentage with other already reported OR/GR gene sequences (Robertson et al.,
592 2019), confirming the conserved protein domain (Sayers et al., 2022). Once we annotated the
593 candidate OR/GR genes for each aphid species, we launched a second round of gene detection
594 using InsectOR (Karpe et al., 2021) to improve the final set of OR and GR gene annotations. As
595 a final validation, we aligned the protein sequences for each gene family using MAFFT v7.4
596 (Kuraku et al., 2013; Katoh et al., 2019) with the standard parameters and a gap-opening
597 penalty of 2. We split or eliminated the extremes in short or highly divergent sequence lengths
598 compared with protein sequences of GRs and ORs, using as references available sequences of
599 ORs and GRs in *A. pisum* and *A. glycines* already reported by Robertson et al. (2019).

600 We performed Spearman rank correlation tests between OR or GR repertoires and
601 genome size estimated from the assemblies. We statistically tested for differences in numbers
602 of OR and GR genes among aphid species using a generalized linear model (glm. nb function)
603 as follows:

604

605 Gene count_{ijk} = μ + species_i + families + ε_{ijk} (Model 1)

606

607 Where 'Gene count_{ijk}' is the number of genes of the species_i from the OR or GR family
608 _j, ' μ ' is the overall mean, species and families are fixed effects, and ' ε_{ijk} ' is the residual. The
609 residual was fitted to a Negative Binomial according to the R package Mass, version 7.3-58.3
610 (Venables and Ripley, 2002R).

611 Annotated GR and OR genes were added to the genome annotation files (gff) retrieved
612 before for the 12 aphid and one-aphid-like species. We used BCFTools 1.5v software to
613 eliminate overlapping regions (Danecek et al., 2021). These new files were the input to detect
614 the orthogroups and for the synteny analysis described below.

615

616 **Ortholog identification and species tree inferences**

617 We used the combined set of predicted proteins for the 13 species to infer orthogroups
618 using OrthoFinder v2.4 (Emms and Kelly, 2019). We then assumed a phylogenetic tree of the
619 13 species using a single-copy orthologous group detected with OrthoFinder v2.4 (Emms and
620 Kelly, 2019) for the 13 genomes. We aligned and concatenated the single-copy orthologs with
621 MAFFT (Kuraku et al., 2013; Katoh et al., 2019). We used TrimAl (Capella-Gutiérrez et al., 2009)
622 to remove the poorly aligned regions with a conservation threshold of 60% and a gap
623 threshold of 90%. We built a maximum likelihood phylogenetic tree with IQ-TREE v2.0.6 (Minh
624 et al., 2020) from the aligned single-copy orthologous sequences and performed 10,000
625 bootstrap replicates. We edited the phylogenetic tree with FigTree v1.4.4
626 (<http://tree.bio.ed.ac.uk/software/figtree/>) with *D. vitifoliae* as an outgroup.

627

628 **Expansion and contraction of gene families**

629 We used the computational analysis of gene family evolution (CAFE v5; Mendes et al.,
630 2020) to detect rapidly evolving gene families, which depict a fast rate of genomic turnover
631 (gains and losses per gene per million years). CAFE estimates the birth-death parameter λ
632 (probability that any gene will be gained or lost at each node and terminal branch, accounting
633 for gene family expansions and contractions) along a provided phylogenetic tree scaled to
634 time with the same length of root tips and gene family count estimated with OrthoFinder
635 (Emms and Kelly, 2019).

636 First, we built an ultrametric species tree from the IQ-TREE mentioned above using r8s
637 software (<http://sourceforge.net/projects/r8s/>). We used as inputs the divergence time and a
638 reference species tree built with the TimeTree website (www.timetree.org) and five available

639 time references (Kim et al., 2011; Ren et al., 2013; Hardy et al., 2015).

640 We ran the CAFE software using the ultrametric tree and gene family counting inferred
641 with Orthofinder. Using a Python script, we removed the gene families that presented gene
642 copies in only one species and with more than 100 gene copies in one or more species. This
643 correction on gene counting inferred by OrthoFinder avoids potential errors in evolutionary
644 rate estimations (Mendes et al., 2020; https://github.com/hahnlab/cafe_tutorial/). We ran
645 CAFE with several modes in which the net gain and loss rate are estimated as 1) a single (λ)
646 and 2) multiple parameters for each gene family over the whole phylogeny. We defined a
647 single evolutionary rate (gene birth-death rate) as a null hypothesis along the different
648 orthogroups and, as an alternative hypothesis, different evolutionary rates on each branch in
649 the phylogeny. The evolutionary hypothesis consisted of testing the host preference (tree
650 versus herbaceous host), specialization (generalist vs. specialist), phylogeny, and taxonomy
651 rank (family, subfamily, tribe, and genus). Finally, we determined the better-fit model between
652 the alternative and null models, detecting significant differences in their log-likelihoods
653 distribution after launching a likelihood-ratio test (LRT; Magis et al., 2010; Thissen et al., 2013)
654 script on RStudio v1.3.1.

655 We launched a Gene Ontology (GO) enrichment analysis using the 283 expanded and
656 100 contracted group genes detected by CAFE. We used the TopGo package (Alexa et al.,
657 2006) to identify the detected genes relevant to biological processes or functions based on
658 GO and the elim method to understand the evolutionary history of gene families.

659

660 **Selective pressures acting on OR and GR genes**

661 We inferred selection within subfamilies defined from the phylogenetic trees of OR and
662 GR genes we built. OR and GR subfamilies were monophyletic groups supported by high
663 bootstrap values (>80%). We also visually inspected the tree topology. We then defined the
664 foreground and background branches for the selection analysis.

665 First, we built two phylogenetic trees for GR and OR genes. Fructose and Orco
666 genes were used as outgroups (Robertson et al., 2019) to root the trees. From the protein files

667 obtained from each GR or OR gene, we retrieved their coding (CDS) sequence from the original
668 fasta file of the genome assembly of the respective species. We aligned the sequences using
669 MUSCLE (Edgar, 2004) with the default parameters and a maximum number of iterations of
670 16. We used Gblocks (Castresana, 2000) to curate the sequences, limiting contiguous non-
671 conserved positions. We also eliminated lines that introduced significant gaps in the
672 alignment. We converted the sequences to codon alignment with PAL2NAL (Suyama et al.,
673 2006). We built the phylogenetic trees (separately for OR and GR genes, respectively) with IQ-
674 TREE v2.0 (Minh et al., 2020) with the following parameters: model finder option, ultrafast
675 bootstrap 1,000, and branch support test (SH-aLRT) 1,000. OR and GR subfamilies were visually
676 defined on the phylogenetic trees as monophyletic groups supported by a bootstrap >80%.

677 We used the maximum likelihood approach implemented in codon-based analysis
678 (*codeml*) implemented in PAML version 4.2 (Yang, 2007) to infer evidence of positive and
679 purifying selection for each GR and OR subfamily. Codeml computes the likelihood of a large
680 set of molecular evolution models that estimate ω , the ratio of the non-synonymous mutation
681 rate (dN) over the synonymous mutation rate (dS) (Kimura, 1983). The most likely model of
682 molecular evolution explaining the data is chosen based on a likelihood ratio test (LRT) statistic
683 $2\Delta L$ (with a χ^2 distribution with degrees of freedom equal to the differences in the number of
684 parameters) computed among models. Once the most likely model is chosen, evidence of
685 positive selection is inferred from the estimates of ω : $\omega > 1$ indicates positive selection, while
686 $\omega < 1$ indicates purifying selection.

687 We utilized the branch (free-ratio) and the site models. We did not use the branch-site
688 model as we neither had any phylogenetic background nor an apparent chemosensory gene
689 family clustering to test for selection. The branch model assumes different ω ratios for different
690 branches on the phylogeny (Yang 1998; Yang and Nielsen 1998) and is used to detect positive
691 selection acting on particular sequences of the subfamilies without averaging ω throughout
692 the phylogenetic tree. Using a Likelihood Ratio Test, the branch model is compared to a null
693 hypothesis ($\omega=1$, one-ratio M0; Nielsen and Yang, 1998).

694 We also tested the site models that can be used to identify positively selected sites in
695 a multiple-sequence alignment (Yang and Nielsen, 2002). The models can be site-class-

696 specific, all of which assume that the ω ratio is the same across branches of the phylogeny but
697 different among sites in the alignment. These codon substitution models are M0 (one-ratio),
698 M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), M8 (beta and $\omega > 1$),
699 and M8a (beta and $\omega = 1$). We applied the LTR statistics to evaluate the fit of the nested models
700 to the data for positive selection (Nielsen and Yang, 1998) confronting M0 vs. M3, M1a vs.
701 M2a, and M7 vs. M8 models. Support for positive selection can be identified if M2a provides
702 a better fit than M1a or M8 provides a better fit than M7 or M8a (Yang et al., 2000; Anisimova
703 et al., 2001). We confirmed signals of positive selection when a better fit of the selected model
704 was supporting M2a and M8.

705 The dN/dS ratio estimated from the branch models for each gene of each subfamily of
706 each gene family (OR, GR, and the Orthologous Groups or OG) of each species was fitted to
707 three statistical models to compare the rate of evolution of ORs, GRs, and OGs. First dN/dS
708 ratio was fitted to a negative binomial distribution (function *glm.nb* in R package lme4) as
709 follows :

$$710 \quad dN/dS_{ijkl} = \mu + \text{species}_i + \text{subfamily}_j + \text{family}_k + \varepsilon_{ijkl} \text{ (Model 1).}$$

711

712 Where ' dN/dS_{ijk} ' is the dN/dS of the sequence from subfamily j from the OR/GR/OG
713 family k of the species i , ' μ ' is the overall mean, species and families are fixed effects, and ' ε_{ijk} '
714 is the residual. The residual was fitted to a Negative Binomial using the R package Mass, version
715 7.3-58.3 (Ripley et al., 2013). Model 2 was like Model 1 but removed the subfamily effect. In
716 model 3, the dN/dS ratio was fitted to a negative binomial distribution (function *glm.nb* in R
717 package lme4) as follows, but assuming the subfamily as a random effect :

$$718 \quad dN/dS_{ijkl} = \mu + \text{species}_i + (1|\text{subfamily})_j + \text{family}_k + \varepsilon_{ijkl} \text{ (Model 3).}$$

719 We compared the fit of three models to the data using the *anova* function (aov R
720 package, Chambers et al., 1992). The statistical significance of an effect was assessed with an
721 ANOVA-type test and Tukey's Honest Significance Test (R function Tukey-HSD) to compare
722 dN/dS among gene families (OG, GR, OR) pairwise. Predicted dN/dS values from the best

723 were then plotted with the `plot_model` function (sjPlot R package, Lüdecke, 2018).

724 **Micro-synteny analyses**

725 Expanded gene families often occur as tandem arrays, a genomic architecture that can
726 contribute to increased gene birth and death rates, increasing copy number variation among
727 species (Ohno 1970). Therefore, we examined how genomic organization varies between GR
728 and OR subfamilies in aphid species to generate insights into the molecular evolutionary
729 mechanisms shaping OR and GR subfamily function. We built a synteny network using the
730 R/Bioconductor package `syntenet` (Almeida-Silva, et al., 2022). We used each species
731 chemosensory protein sequences and the genomic coordinates (gff files). We explored the
732 synteny relationships defining clusters containing only the chemosensory genes. We visualized
733 the phylogenomic profiles of the synteny genes as a heatmap. We also tested for whole-
734 genome duplications (WGD) and small-scale duplications (SSD) using the
735 Rpackage/bioconductor `doubletrouble` package (Almeida and Van de Peer, 2023). We
736 performed the protein alignment using DIAMOND (Buchfink et al., 2015) with the sensitive
737 mode ($e\text{-value} = 1e\text{-}10$, top hits = 5) to detect the WGD and SSD paralogous genes based on
738 the syntenic blocks.

739

740 **Repetitive DNA element annotation**

741 We used the REPET package light version (Quesneville et al., 2005; Flutre et al., 2011)
742 for *de novo* identification and annotation of TEs of the 13 genomes that consisted of utilizing
743 a single clustering software (GROUPER) instead of the three conventional REPET pipeline
744 (GROUPER, RECON, and PILER). REPET includes two main pipelines with a qualitative and
745 quantitative assessment, respectively. We utilized the TEs consensus to represent each type of
746 repetitive element inferred from the TEdenovo pipeline (Flutre et al. 2011) for the TE. We built
747 consensus sequences only containing at least three similar copies identified in the genome
748 using PASTEC (Hoede et al., 2014). We used the TEannot pipeline (Quesneville et al. 2005) to
749 annotate all repetitive elements in the genome utilizing the TEs consensus library and to
750 construct a library in which redundant consensus were eliminated (length $\geq 98\%$, identity

751 $\geq 95\%$). This library of TEs was finally used to annotate the aphids genomes with the TEannot
752 pipeline. First, TEdenovo detects repeated elements in a genome and builds a consensus
753 database. TEdenovo uses BLASTER (Quesneville et al., 2003) to compare each genome with
754 itself. Second, TEannot uses the consensus database to count all the TE copies in the genome.
755 The final product of this first stage of the REPET pipeline was a library of classified and non-
756 redundant consensus sequences. We used the TEs sequences library from TEdenovo to get
757 consensus sequences based on similarity using BLASTER, RepeatMasker (Smit et al., 2015), and
758 CENSOR (Jurka et al., 2005). We annotated the short, simple repeats (SSRs) using TRF (Benson,
759 1999) and RepeatMasker (Smit et al., 2015) and removed them. We applied the TEs joining
760 fragment pipeline (chain pipeline) using MATCHER (Quesneville et al., 2003, 2005). Finally, we
761 used the lengthy join procedure to connect distant fragments (Ahmed et al., 2011).

762

763 We test whether the TE content of each order among aphid species using a general
764 linear model (*glm*) with a quasipoisson distribution and ANOVA test with R/Rstudio
765 (<http://cran.r-project.org/mirrors.html>) as follows:

766

767 $\text{TE count}_{ijk} = \text{Species}_i + \text{TE order}_j + \varepsilon_{ijk}$ (Model 4).

768

769 We included the number of TE copies detected per family (TEs count), the analyzed
770 species (Species), and the type of the TEs order. We also retrieved the cluster of the TE copy
771 identified by REPET to prepare the consensus (Flutre et al., 2011). We compared the TE-copy
772 identity among species using an upset plot (UpSetR R package, Conway et al., 2017).

773

774 4.1.1 TE family divergence

775 We evaluated the TE's contribution the aphid and aphid-like genome due to the significant
776 variation of TE content. We calculated the Kimura distances (sequence divergence) using scripts

777 of the RepeatMasker package (Flynn et al., 2019). We used the RepeatMasker annotation as input
778 to generate the aphid species TEs landscape. Kimura's two-parameter (*K2P*) model is a nucleotide
779 substitution model for estimating genetic distances and phylogenetic relationships, considering
780 transitional and transversional substitution rates (Kimura, 1980). We plot the proportion of base
781 pair repeats (Y-axis) at different levels of divergence (X-axis) calculated as a $2-p$ Kimura distance
782 using the ggplot2 R package (Wickham, 2016). We also utilized the ggplot2 R package to create
783 a plot describing the aphid TE content distribution using the Kimura substitution level as a
784 function of the percentage of the genome. We defined the TE content according to Wicker's TE
785 classification (Wicker et al., 2007) utilized in the PASTEC classifier pipeline from the REPET package
786 (Quesneville et al., 2005). We defined five orders for class I retrotransposons: LTR, DIRS, PLE, LINE,
787 and SINE. We represented the class II TEs or DNA transposons by four orders: TIR, Crypton,
788 Helitron, and Maverick. We observed additional categories of non-autonomous TEs: LARD, TRIM,
789 and MITE.

790

791 **Demographic history analysis**

792 We investigated the demographic histories of the different aphid species with whole-
793 genome datasets (except the *Aphis glycines* genome, which has too low coverage, and *Aphis*
794 *gossypii*, for which no short reads data were available, Table S1) using the pairwise sequentially
795 Markovian coalescent (PSMC) approach (Li and Durbin, 2011), which can infer changes in
796 effective population size (N_e) within 10 000 to 1 000 000 years. We performed PSMC (Li and
797 Durbin, 2011) with the heterozygous SNP loci produced by the BWA-MEM/Samtools pipeline
798 (Li et al., 2009). Parameters were set to "N25 -t15 -r5 -p 4+25*2+4+6" and bootstrapping (100
799 times) was performed for each species to determine the variance in N_e estimates. After
800 obtaining the PSMC output, the previously published (Mathers et al., 2023) mutation rate of
801 $2.7e-10$ per nucleotide per generation and a generation time of 1 year were used to generate
802 the scaled plot using the psmc_plot.pl script. For bootstrapping analyses, the psmcfa file was
803 first split into equal lengths of 0.5 MB and was used for 100 runs of PSMC.

804 One repeat class was unmasked at a time, keeping all the other repeat classes masked

805 before PSMC analyses to evaluate the effect of each repeat class on the PSMC inference.

806

807 **TE enrichment and depletion in the vicinity of OR and GR genes**

808 We analyzed the genomic architecture of OR/GR genes. We used the locus overlap
809 analysis within the LOLA R package Bioconductor (Sheffield and Bock, 2016) to test for TE
810 enrichment and depletion near OR/GR genes in the 12 aphids and the one aphid-like genomes.
811 We created three different datasets to launch the analysis. The query set contained genomic
812 regions of 10 Kb around each OR/GR gene. The region universe (or the control genomic
813 region) comprised gene coordinates of all genic areas of the genome and expanded 10 Kb
814 around each gene. We extended this region because the chemosensory genes used to be
815 organized in clusters within the genome. The reference dataset contained the TE copy
816 coordinates from the REPET analysis. We also repeated the same procedure but restringing
817 the region to 2Kb. Promoters were suggested in 2Kb from the start codon in the gene
818 architecture (Liu and Zhang, 2002; Simola et al., 2013; Chen et al., 2020; Liu et al., 2021); we,
819 therefore, test to find TEs related to gene regulation. We tested for TE enrichment and
820 depletion within OR/GR regions compared to the control regions using the Fisher's Exact Test
821 with false discovery rate correction (q -value<0.1).

822

823 **Data availability**

824 The newly assembled genome of *D. plantaginea* is publicly available on the BIPAA
825 platform (https://bipaa.genouest.org/sp/dysaphis_plantaginea/) and NCBI (XXXX). The OR/GR
826 sequence databases are available SRAXXXX. TE annotations are also available on NCBI
827 (SRAXXXX). All analysis files and codes are available at Forgemia
828 (https://forgemia.inra.fr/sergio.olvera-vazquez/aphid_chemosensory_genes).

829 **Acknowledgments**

830 We appreciate the support of Alexandre Degrave (Université Angers, Institut Agro,
831 INRAE, IRHS, SFR QUASAV), Claire Mottet, Elorri Segura (Unité Résistance aux Produits

832 Phytosanitaires, Laboratoire de Lyon, ANSES), and Christelle Buchard (Institut de Génétique,
833 Environnement et Protection des Plantes, INRAE, Institut Agro, Université Rennes) to get the
834 *D. plantaginea* individuals for the *de novo* assembly. We thank the Bioinformatics Platform for
835 Agroecosystem Arthropods (BIPAA) platform, particularly Fabrice Legeai (Institut de
836 Génétique, Environnement et Protection des Plantes, INRAE, Institut Agro, Université Rennes)
837 and the Centre National de Ressources Génomiques Végétales-CNRGV, especially William
838 Marande, Nathalie Rodde, and Stephane Cauet (INRAE) in the *de novo* assembly of *D.*
839 *plantaginea*. We thank the Plant Bioinformatics Facility, BioinfOmics, INRAE, Université Paris-
840 Saclay, especially Johann Confais and Nathalie Choisne, for their assistance in the TEs
841 annotation of the analyzed species. We thank Hugh David Loxdale for the discussion on
842 specialization in aphids.

843

844 Authors' contributions. AC obtained funding and designed the experiment. CB, DO, AD,
845 ES, BB, FL, and AC prepare the material for genome assemblies; SGOV, AC, XC, AM, FAS, JC, CL,
846 IE, CM, EJJ, and YB analyzed the data; All co-authors discussed the results. The manuscript was
847 written by SGOV, AC, with critical inputs from other co-authors.

848

849 **Funding**

850 ATIP-CNRS, IDEEV, and LabEx BASC funded this research.

851 **Conflict of interest**

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

854

855

856

References

- Abrahamson, W. G. (2008). *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects*. Univ of California Press. <https://doi.org/10.1093/nar/gkr324>
- Abdullaev, E. T., Umarova, I. R., & Arndt, P. F. (2021). Modelling segmental duplications in the human genome. *BMC genomics*, 22, 1-16. <https://doi.org/10.1186/s12864-021-07789-7>
- Ahmad, A., Wallau, G. L., & Ren, Z. (2021). Characterization of Mariner transposons in seven species of Rhus gall aphids. *Scientific reports*, 11(1), 1-11. <https://doi.org/10.1038/s41598-021-95843-5>
- Ahmed, I., Sarazin, A., Bowler, C., Colot, V., & Quesneville, H. (2011). Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. *Nucleic acids research*, 39(16), 6919-6931.
- Alexa, A., Rahnenführer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, 22(13), 1600-1607. <https://doi.org/10.1093/bioinformatics/btl140>
- Almeida, F. C., Sanchez-Gracia, A., Campos, J. L., & Rozas, J. (2014). Family size evolution in Drosophila chemosensory gene families: a comparative analysis with a critical appraisal of methods. *Genome Biology and Evolution*, 6(7), 1669-1682. <https://doi.org/10.1093/gbe/evu130>
- Almeida-Silva, F., Zhao, T., Ullrich, K. K., Schranz, M. E., & Van de Peer, Y. (2023). syntenet: an R/Bioconductor package for the inference and analysis of synteny networks. *Bioinformatics*, 39(1), btac806. <https://doi.org/10.1093/bioinformatics/btac806>
- Almeida-Silva, F., & Van de Peer, Y. (2023). Whole-genome duplications and the long-term evolution of gene regulatory networks in angiosperms. *bioRxiv*, 2023-03. <https://doi.org/10.1101/2023.03.13.532351>
- Amoutzias, G., & Van de Peer, Y. (2008). Together we stand: genes cluster to coordinate regulation. *Developmental cell*, 14(5), 640-642. <https://doi.org/10.1016/j.devcel.2008.04.006>
- Andersson, M. N., Löfstedt, C., & Newcomb, R. D. (2015). Insect olfaction and the evolution of receptor tuning. *Frontiers in Ecology and Evolution*, 3, 53. <https://doi.org/10.3389/fevo.2015.00053>
- Anisimova, M., Bielawski, J. P., & Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular biology and evolution*, 18(8), 1585-1592. <https://doi.org/10.1093/oxfordjournals.molbev.a003945>
- Arnab, S. P., Amin, M. R., & DeGiorgio, M. (2023). Uncovering footprints of natural selection through spectral analysis of genomic summary statistics. *Molecular Biology and Evolution*,

msad157. <https://doi.org/10.1093/molbev/msad157>

Assis, R., & Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in *Drosophila*. *Proceedings of the National Academy of Sciences*, *110*(43), 17409-17414. <https://doi.org/10.1073/pnas.1313759110>

Atamian, H. S., Chaudhary, R., Cin, V. D., Bao, E., Girke, T., & Kaloshian, I. (2013). In planta expression or delivery of potato aphid *Macrosiphum euphorbiae* effectors Me10 and Me23 enhances aphid fecundity. *Molecular Plant-Microbe Interactions*, *26*(1), 67-74. <https://doi.org/10.1094/MPMI-06-12-0144-FI>

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Grothendieck, G. (2009). Package 'lme4'. <http://lme4.r-forge.r-project.org>

Barton, N. H. (1996). Natural selection and random genetic drift as causes of evolution on islands. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1341), 785-795. <https://doi.org/10.1098/rstb.1996.0073>

Ben Amara, W., Quesneville, H., & Khemakhem, M. M. (2021). A genomic survey of *Mayetiola destructor* mobilome provides new insights into the evolutionary history of transposable elements in the cecidomyiid midges. *PLoS one*, *16*(10), e0257996. <https://doi.org/10.1371/journal.pone.0257996>

Benton, R., Sachse, S., Michnick, S. W., & Vosshall, L. B. (2006). Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. *PLoS biology*, *4*(2), e20. <https://doi.org/10.1371/journal.pbio.0040020>

Benton, R. (2015). Multigene family evolution: perspectives from insect chemoreceptors. *Trends in ecology & evolution*, *30*(10), 590-600. <https://doi.org/10.1016/j.tree.2015.07.009>

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., ... & Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution*, *69*(2), 313-319. <https://doi.org/10.1016/j.ympev.2012.08.023>

Best, A., White, A., Kisdi, E., Antonovics, J., Brockhurst, M. A., & Boots, M. (2010). The evolution of host-parasite range. *The American Naturalist*, *176*(1), 63-71. <https://doi.org/10.1086/653002>

Biello, R., Singh, A., Godfrey, C. J., Fernández, F. F., Mugford, S. T., Powell, G., ... & Mathers, T. C. (2021). A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Molecular ecology resources*, *21*(1), 316-326. <https://doi.org/10.1111/1755-0998.13258>

Blackman, R. L., & Eastop, V. F. (2000). *Aphids on the world's crops: an identification and information guide* (No. Ed. 2). John Wiley & Sons Ltd.

Blackman, R. L., & Eastop, V. F. (2008). Aphids on the world's herbaceous plants and shrubs, 2 volume set. John Wiley & Sons.

Bouallègue, M., Filée, J., Kharrat, I., Mezghani-Khemakhem, M., Rouault, J. D., Makni, M., & Capy, P. (2017). Diversity and evolution of mariner-like elements in aphid genomes. *BMC genomics*, *18*(1), 1-12. <https://doi.org/10.1186/s12864-017-3856-6>

Brand, P., Ramírez, S. R., Leese, F., Quezada-Euan, J. J. G., Tollrian, R., & Eltz, T. (2015). Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). *BMC evolutionary biology*, *15*(1), 1-16. <https://doi.org/10.1186/s12862-015-0451-9>

Brand, P., Robertson, H. M., Lin, W., Pothula, R., Klingeman, W. E., Jurat-Fuentes, J. L., & Johnson, B. R. (2018). The origin of the odorant receptor gene family in insects. *Elife*, *7*, e38340. <https://doi.org/10.7554/eLife.38340>

Brockhurst, M. A., Chapman, T., King, K. C., Mank, J. E., Paterson, S., & Hurst, G. D. (2014). Running with the Red Queen: the role of biotic conflicts in evolution. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), 20141382. <https://doi.org/10.1098/rspb.2014.1382>

Bourgeois, Y., & Boissinot, S. (2019). On the population dynamics of junk: a review on the population genomics of transposable elements. *Genes*, *10*(6), 419. <https://doi.org/10.3390/genes10060419>

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature methods*, *12*(1), 59-60. <https://doi.org/10.1038/nmeth.3176>

Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Todd Hubisz, M., Glanowski, S., ... & Clark, A. G. (2005). Natural selection on protein-coding genes in the human genome. *Nature*, *437*(7062), 1153-1157. <https://doi.org/10.1038/nature04240>

Byrne, S., Schughart, M., Carolan, J. C., Gaffney, M., Thorpe, P., Malloch, G., ... & McNamara, L. (2022). Genome sequence of the English grain aphid, *Sitobion avenae* and its endosymbiont *Buchnera aphidicola*. *G3*, *12*(3), jkab418. <https://doi.org/10.1093/g3journal/jkab418>

Cacciò, S. M., Chalmers, R. M., Dorny, P., & Robertson, L. J. (2018). Foodborne parasites: Outbreaks and outbreak investigations. A meeting report from the European network for foodborne parasites (Euro-FBP). *Food and Waterborne Parasitology*, *10*, 1-5. <https://doi.org/10.1016/j.fawpar.2018.01.001>

Capaul, M., & Ebert, D. (2003). Parasite-mediated selection in experimental *Daphnia magna* populations. *Evolution*, *57*(2), 249-260. <https://doi.org/10.1111/j.0014-3820.2003.tb00260.x>

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972-1973.

<https://doi.org/10.1093/bioinformatics/btp348>

Casacuberta, E., & González, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular ecology*, 22(6), 1503-1517. <https://doi.org/10.1111/mec.12170>

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, 17(4), 540-552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>

Chen, W., Shakir, S., Bigham, M., Richter, A., Fei, Z., & Jander, G. (2019). Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). *Gigascience*, 8(4), giz033. <https://doi.org/10.1093/gigascience/giz033>

Chen, X., Tan, A., & Palli, S. R. (2020). Identification and functional analysis of promoters of heat-shock genes from the fall armyworm, *Spodoptera frugiperda*. *Scientific reports*, 10(1), 2363 <https://doi.org/10.1038/s41598-020-59197-8>

Chénais, B., Caruso, A., Hiard, S., & Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene*, 509(1), 7-15. <https://doi.org/10.1016/j.gene.2012.07.042>

Choi, H., Shin, S., Jung, S., Clarke, D. J., & Lee, S. (2018). Molecular phylogeny of Macrosiphini (Hemiptera: Aphididae): An evolutionary hypothesis for the Pterocomma-group habitat adaptation. *Molecular Phylogenetics and Evolution*, 121, 12-22. <https://doi.org/10.1016/j.ympev.2017.12.021>

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*, 18(2), 71-86. <https://doi.org/10.1038/nrg.2016.139>

Chyb, S. (2004). *Drosophila* gustatory receptors: from gene identification to functional expression. *Journal of insect physiology*, 50(6), 469-477. <https://doi.org/10.1016/j.jinsphys.2004.03.012>

Clay, K., & Kover, P. X. (1996). The Red Queen hypothesis and plant/pathogen interactions. *Annual review of Phytopathology*, 34(1), 29-50. <https://doi.org/10.1146/annurev.phyto.34.1.29>

Clyne, P. J., Warr, C. G., Freeman, M. R., Lessing, D., Kim, J., & Carlson, J. R. (1999). A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron*, 22(2), 327-338. [https://doi.org/10.1016/S0896-6273\(00\)81093-4](https://doi.org/10.1016/S0896-6273(00)81093-4)

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33(18), 2938-2940. <https://doi.org/10.1093/bioinformatics/btx364>

- Culver, D. C., Kowalko, J. E., & Pipan, T. (2023). Natural selection versus neutral mutation in the evolution of subterranean life: A false dichotomy?. *Frontiers in Ecology and Evolution*, *11*, 1080503. <https://doi.org/10.3389/fevo.2023.1080503>
- De Bruyne, M., & Baker, T. C. (2008). Odor detection in insects: volatile codes. *Journal of chemical ecology*, *34*(7), 882-897. <https://doi.org/10.1007/s10886-008-9485-4>
- Dalton, R. P., & Lomvardas, S. (2015). Chemosensory receptor specificity and regulation. *Annual review of neuroscience*, *38*, 331. <https://doi.org/10.1146/annurev-neuro-071714-034145>
- Dangl, J. L., & Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *nature*, *411*(6839), 826-833. <https://doi.org/10.1038/35081161>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., ... & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- Date, P., Dweck, H. K., Stensmyr, M. C., Shann, J., Hansson, B. S., & Rollmann, S. M. (2013). Divergence in olfactory host plant preference in *D. mojavensis* in response to cactus host use. *PLOS one*, *8*(7), e70027. <https://doi.org/10.1371/journal.pone.0070027>
- Desvignes, T., Lauridsen, H., Valdivieso, A., Fontenele, R. S., Kraberger, S., Murray, K. N., ... & Postlethwait, J. H. (2022). A parasite outbreak in notothenioid fish in an Antarctic fjord. *Iscience*, *25*(7). <https://doi.org/10.1016/j.isci.2022.104588>
- Diepeveen, E. T., Kim, F. D., & Salzburger, W. (2013). Sequence analyses of the distal-less homeobox gene family in East African cichlid fishes reveal signatures of positive selection. *BMC Evolutionary Biology*, *13*(1), 1-12 <https://doi.org/10.1186/1471-2148-13-153>
- Dixon, A. F. G. (1977). Aphid ecology: life cycles, polymorphism, and population regulation. *Annual Review of Ecology and Systematics*, 329-353. <https://www.jstor.org/stable/2096732>
- Dondini, L., Lugli, S., & Sansavini, S. (2018). Cherry breeding: sweet cherry (*Prunus avium* L.) and sour cherry (*Prunus cerasus* L.). In *Advances in Plant Breeding Strategies: Fruits: Volume 3* (pp. 31-88). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-91944-7>
- Duplessis, S., Cuomo, C. A., Lin, Y. C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., ... & Martin, F. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*, *108*(22), 9166-9171. <https://doi.org/10.1073/pnas.1019315108>
- Duran, C., Edwards, D., & Batley, J. (2009). Genetic maps and the use of synteny. *Plant genomics: methods and protocols*, 41-55. https://doi.org/10.1007/978-1-59745-427-8_3

Dybdahl, M. F., & Storfer, A. (2003). Parasite local adaptation: red queen versus suicide king. *Trends in Ecology & Evolution*, 18(10), 523-530. [https://doi.org/10.1016/S0169-5347\(03\)00223-4](https://doi.org/10.1016/S0169-5347(03)00223-4)

Eastop, V.F. Deductions from the Present Day Host Plants of Aphids and Related Insects. Roy Entomol. Soc. Lond. Symp. 1972, 6, 157–178.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), 1792-1797. <https://doi.org/10.1093/nar/gkh340>

Elliott, T. A., & Gregory, T. R. (2015). Do larger genomes contain more diverse transposable elements?. *BMC evolutionary biology*, 15, 1-10. <https://doi.org/10.1186/s12862-015-0339-8>

Emden, H. V., & Harrington, R. (Eds.). (2017). *Aphids as crop pests*. Cabi.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, 20(1), 1-14. <https://doi.org/10.1186/s13059-019-1832-y>

Engsontia, P., Sangket, U., Chotigeat, W., & Satasook, C. (2014). Molecular evolution of the odorant and gustatory receptor genes in lepidopteran insects: implications for their adaptation and speciation. *Journal of Molecular Evolution*, 79, 21-39. <https://doi.org/10.1007/s00239-014-9633-0>

Eickbush, T. H., & Furano, A. V. (2002). Fruit flies and humans respond differently to retrotransposons. *Current opinion in genetics & development*, 12(6), 669-674. [https://doi.org/10.1016/S0959-437X\(02\)00359-3](https://doi.org/10.1016/S0959-437X(02)00359-3)

Fay, J. C., & Wittkopp, P. J. (2008). Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*, 100(2), 191-199. <https://doi.org/10.1038/sj.hdy.6801000>

Fernández, R., Marcet-Houben, M., Legeai, F., Richard, G., Robin, S., Wucher, V., ... & Tagu, D. (2020). Selection following gene duplication shapes recent genome evolution in the pea aphid *Acyrtosiphon pisum*. *Molecular biology and evolution*, 37(9), 2601-2615. <https://doi.org/10.1093/molbev/msaa110>

Ferry, N., Edwards, M. G., Gatehouse, J. A., & Gatehouse, A. M. (2004). Plant–insect interactions: molecular approaches to insect resistance. *Current opinion in Biotechnology*, 15(2), 155-161. <https://doi.org/10.1016/j.copbio.2004.01.008>

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*, 9(5), 397-405. <https://doi.org/10.1038/nrg2337>

Filée, J., Rouault, J. D., Harry, M., & Hua-Van, A. (2015). Mariner transposons are sailing in the genome of the blood-sucking bug *Rhodnius prolixus*. *BMC genomics*, 16(1), 1-17. <https://doi.org/10.1186/s12864-015-2060-9>

Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit. 2019. "RepeatModeler2: Automated Genomic Discovery of Transposable Element Families." Preprint. Genomics. <https://doi.org/10.1101/856591>

Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PloS one*, 6(1), e16526. <https://doi.org/10.1371/journal.pone.0016526>

Fouché, S., Badet, T., Oggenfuss, U., Plissonneau, C., Francisco, C. S., & Croll, D. (2020). Stress-driven transposable element de-repression dynamics and virulence evolution in a fungal pathogen. *Molecular Biology and Evolution*, 37(1), 221-239. <https://doi.org/10.1093/molbev/msz216>

Frantzeskakis, L., Kusch, S., & Panstruga, R. (2019). The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Molecular Plant Pathology*, 20(1), 3. <https://doi.org/10.1111/mpp.12738>

Frantzeskakis, L., Di Pietro, A., Rep, M., Schirawski, J., Wu, C. H., & Panstruga, R. (2020). Rapid evolution in plant-microbe interactions—a molecular genomics perspective. *New Phytologist*, 225(3), 1134-1142. <https://doi.org/10.1111/nph.15966>

Futuyma, D. J., & Agrawal, A. A. (2009). Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences*, 106(43), 18054-18061. <https://doi.org/10.1073/pnas.0904106106>

Gardiner, A., Barker, D., Butlin, R. K., Jordan, W. C., & Ritchie, M. G. (2008). Drosophila chemoreceptor gene evolution: selection, specialization and genome size. *Molecular Ecology*, 17(7), 1648-1657. <https://doi.org/10.1111/j.1365-294X.2008.03713.x>

Gilbert, C., Peccoud, J., & Cordaux, R. (2021). Transposable elements and the evolution of insects. *Annual Review of Entomology*, 66(1), 355-372. <https://doi.org/10.1146/annurev-ento-070720-074650>

Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguilera, G., De Vienne, D. M., ... & Giraud, T. (2014). Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular ecology*, 23(4), 753-773. <https://doi.org/10.1111/mec.12631>

Gomes, G. B., Hutson, K. S., Domingos, J. A., Chung, C., Hayward, S., Miller, T. L., & Jerry, D. R. (2017). Use of environmental DNA (eDNA) and water quality data to predict protozoan parasites outbreaks in fish farms. *Aquaculture*, 479, 467-473. <https://doi.org/10.1016/j.aquaculture.2017.06.021>

Goubert, C., Modolo, L., Vieira, C., ValienteMoro, C., Mavingui, P., & Boulesteix, M. (2015). De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome biology and evolution*, 7(4), 1192-1205.

<https://doi.org/10.1093/gbe/evv050>

Gowler, C. D., Essington, H., O'Brien, B., Shaw, C. L., Bilich, R. W., Clay, P. A., & Duffy, M. A. (2023). Virulence evolution during a naturally occurring parasite outbreak. *Evolutionary Ecology*, 37(1), 113-129. <https://doi.org/10.37807/GBMF9202>

Guo, S., & Kim, J. (2007). Molecular evolution of *Drosophila* odorant receptor genes. *Molecular biology and evolution*, 24(5), 1198-1207. <https://doi.org/10.1093/molbev/msm038>

Hall, A. R., Scanlan, P. D., Morgan, A. D., & Buckling, A. (2011). Host–parasite coevolutionary arms races give way to fluctuating selection. *Ecology letters*, 14(7), 635-642. <https://doi.org/10.1111/j.1461-0248.2011.01624.x>

Hahn, M. W., De Bie, T., Stajich, J. E., Nguyen, C., & Cristianini, N. (2005). Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome research*, 15(8), 1153-1160. <https://doi.org/10.1101/gr.3567505>

Hallem, E. A., Dahanukar, A., & Carlson, J. R. (2006). Insect odor and taste receptors. *Annu. Rev. Entomol.*, 51, 113-135 <https://doi.org/10.1146/annurev.ento.51.051705.113646>

Hansson, B. S., & Stensmyr, M. C. (2011). Evolution of insect olfaction. *Neuron*, 72(5), 698-711. <https://doi.org/10.1016/j.neuron.2011.11.003>

Hardy, N. B., & Otto, S. P. (2014). Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proceedings of the Royal Society B: Biological Sciences*, 281(1795), 20132960. <https://doi.org/10.1098/rspb.2013.2960>

Hardy, N. B., Peterson, D. A., & von Dohlen, C. D. (2015). The evolution of life cycle complexity in aphids: Ecological optimization or historical constraint?. *Evolution*, 69(6), 1423-1432. <https://doi.org/10.1111/evo.12643>

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *The ISME journal*, 11(5), 1189-1204. <https://doi.org/10.1038/ismej.2016.196>

He, H., Crabbe, M. J. C., & Ren, Z. (2023). Genome-wide identification and characterization of the chemosensory relative protein genes in *Rhus* gall aphid *Schlechtendalia chinensis*. *BMC genomics*, 24(1), 222. <https://doi.org/10.1186/s12864-023-09322-4>

Hickner, P. V., Rivaldi, C. L., Johnson, C. M., Siddappaji, M., Raster, G. J., & Syed, Z. (2016). The making of a pest: Insights from the evolution of chemosensory receptor families in a pestiferous and invasive fly, *Drosophila suzukii*. *BMC genomics*, 17, 1-17. <https://doi.org/10.1186/s12864-016-2983-9>

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., & Quesneville, H.

(2014). PASTEC: an automatic transposable element classification tool. *PloS one*, 9(5), e91929. <https://doi.org/10.1371/journal.pone.0091929>

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics*, 12(1), 1-14. <https://doi.org/10.1186/1471-2105-12-491>

Huang, T., Liu, Y., He, K., Francis, F., Wang, B., & Wang, G. (2023). Chromosome-level genome assembly of the spotted alfalfa aphid *Therioaphis trifolii*. *Scientific Data*, 10(1), 274. <https://doi.org/10.1038/s41597-023-02179-y>

Huerta-Cepas, J., Marcet-Houben, M., Pignatelli, M., Moya, A., & Gabaldón, T. (2010). The pea aphid phylome: a complete catalogue of evolutionary histories and arthropod orthology and paralogy relationships for *Acyrtosiphon pisum* genes. *Insect Molecular Biology*, 19, 13-21. <https://doi.org/10.1111/j.1365-2583.2009.00947.x>

International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS biology*, 8(2), e1000313. <https://doi.org/10.1371/journal.pbio.1000313>

Jackman, S. D., Coombe, L., Chu, J., Warren, R. L., Vandervalk, B. P., Yeo, S., ... & Birol, I. (2018). Tigmint: correcting assembly errors using linked reads from large molecules. *BMC bioinformatics*, 19(1), 1-10. <https://doi.org/10.1186/s12859-018-2425-6>

Jiang, X. J., Ning, C., Guo, H., Jia, Y. Y., Huang, L. Q., Qu, M. J., & Wang, C. Z. (2015). A gustatory receptor tuned to D-fructose in antennal sensilla chaetica of *Helicoverpa armigera*. *Insect Biochemistry and Molecular Biology*, 60, 39-46. <https://doi.org/10.1016/j.ibmb.2015.03.002>

Jiang, X., Zhang, Q., Qin, Y., Yin, H., Zhang, S., Li, Q., ... & Chen, J. (2019). A chromosome-level draft genome of the grain aphid *Sitobion miscanthi*. *Gigascience*, 8(8), giz101. <https://doi.org/10.1093/gigascience/giz101>

Jiao, W. B., & Schneeberger, K. (2020). Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nature Communications*, 11(1), 989. <https://doi.org/10.1038/s41467-020-14779-y>

Joseph, R. M., & Carlson, J. R. (2015). *Drosophila* chemoreceptors: a molecular interface between the chemical world and the brain. *Trends in Genetics*, 31(12), 683-695. <https://doi.org/10.1016/j.tig.2015.09.005>

Julca, I., Marcet-Houben, M., Cruz, F., Vargas-Chavez, C., Johnston, J. S., Gómez-Garrido, J., ... & Gabaldón, T. (2020). Phylogenomics identifies an ancestral burst of gene duplications predating the diversification of Aphidomorpha. *Molecular Biology and Evolution*, 37(3), 730-756. <https://doi.org/10.1093/molbev/msz261>

Kambere, M. B., & Lane, R. P. (2007). Co-regulation of a large and rapidly evolving repertoire

of odorant receptor genes. *BMC neuroscience*, 8(3), 1-16. <https://doi.org/10.1186/1471-2202-8-S3-S2>

Karlsson, E. K., Kwiatkowski, D. P., & Sabeti, P. C. (2014). Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, 15(6), 379-393. <https://doi.org/10.1038/nrg3734>

Karpe, S. D., Tiwari, V., & Ramanathan, S. (2021). InsectOR—Webserver for sensitive identification of insect olfactory receptor genes from non-model genomes. *Plos one*, 16(1), e0245324. <https://doi.org/10.1371/journal.pone.0245324>

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, 20(4), 1160-1166. <https://doi.org/10.1093/bib/bbx108>

Kazazian Jr, H. H. (2004). Mobile elements: drivers of genome evolution. *science*, 303(5664), 1626-1632. <https://doi.org/10.1126/science.10896>

Keller, O., Odrionitz, F., Stanke, M., Kollmar, M., & Waack, S. (2008). Scipio: using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC bioinformatics*, 9(1), 1-12. <https://doi.org/10.1186/1471-2105-9-278>

Kettles, G. J., & Kaloshian, I. (2016). The potato aphid salivary effector Me47 is a glutathione-S-transferase involved in modifying plant responses to aphid infestation. *Frontiers in Plant Science*, 7, 1142. <https://doi.org/10.3389/fpls.2016.01142>

Kim, H., Lee, S., & Jang, Y. (2011). Macroevolutionary patterns in the Aphidini aphids (Hemiptera: Aphididae): diversification, host association, and biogeographic origins. *PLoS One*, 6(9), e24749. <https://doi.org/10.1371/journal.pone.0024749>

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, 16(2), 111-120. <https://doi.org/10.1007/BF01731581>

Kirk, H., Dorn, S., & Mazzi, D. (2013). Molecular genetics and genomics generate new insights into invertebrate pest invasions. *Evolutionary Applications*, 6(5), 842-856. <https://doi.org/10.1111/eva.12071>

Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, M. A., Clark, J. M., ... & Pittendrigh, B. R. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences*, 107(27), 12168-12173. <https://doi.org/10.1073/pnas.1003379107>

Klai, K., Chenais, B., Zidi, M., Djebbi, S., Caruso, A., Denis, F., ... & Mezghani Khemakhem, M. (2020). Screening of *Helicoverpa armigera* mobilome revealed transposable element insertions in insecticide resistance genes. *Insects*, 11(12), 879. <https://doi.org/10.3390/insects11120879>

Koeck, M., Hardham, A. R., & Dodds, P. N. (2011). The role of effectors of biotrophic and hemibiotrophic fungi in infection. *Cellular microbiology*, 13(12), 1849-1857. <https://doi.org/10.1111/j.1462-5822.2011.01665.x>

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology*, 37(5), 540-546. <https://doi.org/10.1038/s41587-019-0072-8>

Kulmuni, J., Wurm, Y., & Pamilo, P. (2013). Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. *Heredity*, 110(6), 538-547. <https://doi.org/10.1038/hdy.2012.122>

Kuraku, S., Zmasek, C. M., Nishimura, O., & Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic acids research*, 41(W1), W22-W28. <https://doi.org/10.1093/nar/gkt389>

Laslett, D., & Canbäck, B. (2008). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*, 24(2), 172-175. <https://doi.org/10.1093/bioinformatics/btm573>

Leal, W. S. (2013). Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annual review of entomology*, 58, 373-391. <https://doi.org/10.1146/annurev-ento-120811-153635>

Leclair, M., Pons, I., Mahéo, F., Morlière, S., Simon, J. C., & Outreman, Y. (2016). Diversity in symbiont consortia in the pea aphid complex is associated with large phenotypic variation in the insect host. *Evolutionary Ecology*, 30(5), 925-941. <https://doi.org/10.1007/s10682-016-9856-1>

Lee, E., Helt, G. A., Reese, J. T., Munoz-Torres, M. C., Childers, C. P., Buels, R. M., ... & Lewis, S. E. (2013). Web Apollo: a web-based genomic annotation editing platform. *Genome biology*, 14(8), 1-13. <https://doi.org/10.1186/gb-2013-14-8-r93>

Lee, S., & Lee, S. (2020). Multigene phylogeny uncovers oviposition-related evolutionary history of Cerambycinae (Coleoptera: Cerambycidae). *Molecular phylogenetics and evolution*, 145, 106707. <https://doi.org/10.1016/j.ympev.2019.106707>

Legeai, F., Shigenobu, S., Gauthier, J. P., Colbourne, J., Rispe, C., Collin, O., ... & Tagu, D. (2010). AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome. *Insect molecular biology*, 19, 5-12. <https://doi.org/10.1111/j.1365-2583.2009.00930.x>

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & 1000 Genome Project

Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *bioinformatics*, 25(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493-496. <https://doi.org/10.1038/nature10231>

Li, Y., Park, H., Smith, T. E., & Moran, N. A. (2019). Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Molecular biology and evolution*, 36(10), 2143-2156. <https://doi.org/10.1093/molbev/msz138>

Li, Z. Q., Luo, Z. X., Cai, X. M., Bian, L., Xin, Z. J., Liu, Y., ... & Chen, Z. M. (2017). Chemosensory gene families in *Ectopis grisescens* and candidates for detection of type-II sex pheromones. *Frontiers in Physiology*, 8, 953. <https://doi.org/10.3389/fphys.2017.00953>

Lively, C. M., & Dybdahl, M. F. (2000). Parasite adaptation to locally common host genotypes. *Nature*, 405(6787), 679-681. <https://doi.org/10.1038/35015069>

Liu, N., & Zhang, L. (2002). Identification of two new cytochrome P450 genes and their 5'-flanking regions from the housefly, *Musca domestica*. *Insect biochemistry and molecular biology*, 32(7), 755-764. <https://doi.org/10.1186/s12915-021-01004-5>

Liu, Y., Liu, H., Wang, H., Huang, T., Liu, B., Yang, B., ... & Wang, G. (2021). *Apolygus lucorum* genome provides insights into omnivorousness and mesophyll feeding. *Molecular ecology resources*, 21(1), 287-300. <https://doi.org/10.1111/1755-0998.13253>

Liu, Q., Jiang, F., Zhang, J., Li, X., & Kang, L. (2021). Transcription initiation of distant core promoters in a large-sized genome of an insect. *BMC biology*, 19, 1-21. <https://doi.org/10.1186/s12915-021-01004-5>

Lockton, S., Ross-Ibarra, J., & Gaut, B. S. (2008). Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences*, 105(37), 13965-13970. <https://doi.org/10.1073/pnas.0804671105>

Loxdale, H. D. (2008). The nature and reality of the aphid clone: genetic variation, adaptation and evolution. *Agricultural and Forest Entomology*, 10(2), 81-90. <https://doi.org/10.1111/j.1461-9563.2008.00364.x>

Lynch, M., & Walsh, B. (2007). *The origins of genome architecture* (Vol. 98). Sunderland, MA: Sinauer associates.

Lynch, M. (2023). Mutation pressure, drift, and the pace of molecular coevolution. *Proceedings of the National Academy of Sciences*, 120(27), e2306741120. <https://doi.org/10.1073/pnas.2306741120>

Ma, L., Li, M. Y., Chang, C. Y., Chen, F. F., Hu, Y., & Liu, X. D. (2019). The host range of *Aphis*

gossypii is dependent on aphid genetic background and feeding experience. *PeerJ*, 7, e7774. <https://doi.org/10.7717/peerj.7774/supp-1>

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of genetics*, 92(1), 155-161. <https://doi.org/10.1007/s12041-013-0212-8>

Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior research methods*, 42(3), 847-862. <https://doi.org/10.3758/BRM.42.3.847>

Mang, D., Shu, M., Endo, H., Yoshizawa, Y., Nagata, S., Kikuta, S., & Sato, R. (2016). Expression of a sugar clade gustatory receptor, BmGr6, in the oral sensory organs, midgut, and central nervous system of larvae of the silkworm *Bombyx mori*. *Insect biochemistry and molecular biology*, 70, 85-98. <https://doi.org/10.1016/j.ibmb.2015.12.008>

Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., Baa-Puyoulet, P., ... & Hogenhout, S. A. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome biology*, 18(1), 1-20. <https://doi.org/10.1186/s13059-016-1145-3>

Mathers, T. C. (2020). Improved genome assembly and annotation of the soybean aphid (*Aphis glycines* Matsumura). *G3: Genes, Genomes, Genetics*, 10(3), 899-906. <https://doi.org/10.1534/g3.119.400954>

Mathers, T. C., Mugford, S. T., Hogenhout, S. A., & Tripathi, L. (2020). Genome sequence of the banana aphid, *Pentalonia nigronervosa* Coquerel (Hemiptera: Aphididae) and its symbionts. *G3: Genes, Genomes, Genetics*, 10(12), 4315-4321. <https://doi.org/10.1534/g3.120.401358>

Mathers, T. C., Wouters, R. H., Mugford, S. T., Swarbreck, D., Van Oosterhout, C., & Hogenhout, S. A. (2021). Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Molecular Biology and Evolution*, 38(3), 856-875. <https://doi.org/10.1093/molbev/msaa246>

McBride, C. S. (2007). Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proceedings of the National Academy of Sciences*, 104(12), 4996-5001. <https://doi.org/10.1073/pnas.0608424104>

McBride, C. S., & Arguello, J. R. (2007). Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics*, 177(3), 1395-1416. <https://doi.org/10.1534/genetics.107.078683>

McCullers, T. J., & Steiniger, M. (2017). Transposable elements in *Drosophila*. *Mobile genetic elements*, 7(3), 1-18. <https://doi.org/10.1080/2159256X.2017.1318201>

McKenzie, S. K., & Kronauer, D. J. (2018). The genomic architecture and molecular evolution of ant odorant receptors. *Genome research*, 28(11), 1757-1765. <http://www.genome.org/cgi/doi/10.1101/gr.237123.118>

Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*, 36(22-23), 5516-5518. <https://doi.org/10.1093/bioinformatics/btaa1022>

Mérel, V., Boulesteix, M., Fablet, M., & Vieira, C. (2020). Transposable elements in *Drosophila*. *Mobile DNA*, 11(1), 1-20. <https://doi.org/10.1186/s13100-020-00213-z>

Meslin, Camille, Pauline Mainet, Nicolas Montagné, Stéphanie Robin, Fabrice Legeai, Anthony Bretaudeau, J Spencer Johnston, et al. 2022. "Spodoptera Littoralis Genome Mining Brings Insights on the Dynamic of Expansion of Gustatory Receptors in Polyphagous Noctuidae." *G3 Genes|Genomes|Genetics*. <https://doi.org/10.1093/g3journal/jkac131>

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37(5), 1530-1534. <https://doi.org/10.1093/molbev/msaa015>

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, 15(12), 756-771. <https://doi.org/10.1038/nrmicro.2017.76>

Moran, N. A. (1992). The evolution of aphid life cycles. *Annual review of entomology*, 37(1), 321-348. <https://doi.org/10.1146/annurev.en.37.010192.001541>

Nicholson, S. J., Nickerson, M. L., Dean, M., Song, Y., Hoyt, P. R., Rhee, H., ... & Puterka, G. J. (2015). The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC genomics*, 16(1), 1-16. <https://doi.org/10.1186/s12864-015-1525-1>

Nielsen, R., & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, 148(3), 929-936. <https://doi.org/10.1093/genetics/148.3.929>

Niimura, Y., & Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS one*, 2(8), e708. <https://doi.org/10.1371/journal.pone.0000708>

Oggenfuss, U., Badet, T., Wicker, T., Hartmann, F. E., Singh, N. K., Abraham, L., ... & Croll, D. (2021). A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *Elife*, 10, e69249. <https://doi.org/10.7554/eLife.69249.sa1>

Olvera-Vazquez, S. G., Remoué, C., Venon, A., Rousselet, A., Grandcolas, O., Azrine, M., ... & Cornille, A. (2021). Large-scale geography survey provides insights into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North

Africa. *Peer Community Journal*, 1. <https://doi.org/10.24072/pcjournal.26>

Papkou, A., Guzella, T., Yang, W., Koepper, S., Pees, B., Schalkowski, R., ... & Schulenburg, H. (2019). The genomic basis of Red Queen dynamics during rapid reciprocal host–pathogen coevolution. *Proceedings of the National Academy of Sciences*, 116(3), 923–928. <https://doi.org/10.1073/pnas.1810402116>

Parisot, N., Vargas-Chávez, C., Goubert, C., Baa-Puyoulet, P., Balmand, S., Beranger, L., ... & Heddi, A. (2021). The transposable element-rich genome of the cereal pest *Sitophilus oryzae*. *BMC biology*, 19(1), 1–28. <https://doi.org/10.1186/s12915-021-01158-2>

Peccoud, Jean, & Simon, Jean-Christophe. (2010). The pea aphid complex as a model of ecological speciation. *Ecological Entomology*, 35, 119–130. <https://doi.org/10.1111/j.1365-2311.2009.01147.x>

Persi, E., Wolf, Y. I., & Koonin, E. V. (2016). Positive and strongly relaxed purifying selection drive the evolution of repeats in proteins. *Nature communications*, 7(1), 13570. <https://doi.org/10.1038/ncomms13570>

Pitino, M., & Hogenhout, S. A. (2013). Aphid protein effectors promote aphid colonization in a plant species-specific manner. *Molecular Plant-Microbe Interactions*, 26(1), 130–139. <https://doi.org/10.1094/MPMI-07-12-0172-FI>

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., & Croll, D. (2017). Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. *Frontiers in plant science*, 8, 119. <https://doi.org/10.3389/fpls.2017.00119>

Plissonneau, C., Hartmann, F. E., & Croll, D. (2018). Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC biology*, 16(1), 1–16. <https://doi.org/10.1186/s12915-017-0457-4>

Podlaha, O., & Zhang, J. (2010). Pseudogenes and their evolution. *eLS*. <https://doi.org/10.1002/9780470015902.a0005118.pub2>

Puinean, A. M., Foster, S. P., Oliphant, L., Denholm, I., Field, L. M., Millar, N. S., ... & Bass, C. (2010). Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLOS genetics*, 6(6), e1000999. <https://doi.org/10.1371/journal.pgen.1000999>

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., ... & Paterson, A. H. (2019). Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome biology*, 20(1), 1–23. <https://doi.org/10.1186/s13059-019-1650-2>

Quan, Q., Hu, X., Pan, B., Zeng, B., Wu, N., Fang, G., ... & Zhan, S. (2019). Draft genome of the cotton aphid *Aphis gossypii*. *Insect Biochemistry and Molecular Biology*, 105, 25–32.

<https://doi.org/10.1016/j.ibmb.2018.12.007>

Quesneville, H., Nouaud, D., & Anxolabéhère, D. (2003). Detection of new transposable element families in *Drosophila melanogaster* and *Anopheles gambiae* genomes. *Journal of molecular evolution*, 57(1), S50-S59. <https://doi.org/10.1007/s00239-003-0007-2>

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., & Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS computational biology*, 1(2), e22. <https://doi.org/10.1371/journal.pcbi.0010022>

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842. <https://doi.org/10.1093/bioinformatics/btq033>

Rabajante, J. F., Tubay, J. M., Ito, H., Uehara, T., Kakishima, S., Morita, S., ... & Ebert, D. (2016). Host-parasite Red Queen dynamics with phase-locked rare genotypes. *Science advances*, 2(3), e1501548. <https://doi.org/10.1126/sciadv.1501548>

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, 10(6), 417-430. <https://doi.org/10.1038/nrmicro2790>

Ragsdale, D. W., Voegtlin, D. J., & O'neil, R. J. (2004). Soybean aphid biology in North America. *Annals of the Entomological Society of America*, 97(2), 204-208. [https://doi.org/10.1603/0013-8746\(2004\)097\[0204:SABINA\]2.0.CO;2](https://doi.org/10.1603/0013-8746(2004)097[0204:SABINA]2.0.CO;2)

Ramdaya, P., & Benton, R. (2010). Evolving olfactory systems on the fly. *Trends in Genetics*, 26(7), 307-316. <https://doi.org/10.1016/j.tig.2010.04.004>

Ren, Z., Zhong, Y., Kurosu, U., Aoki, S., Ma, E., von Dohlen, C. D., & Wen, J. (2013). Historical biogeography of Eastern Asian–Eastern North American disjunct Melaphidina aphids (Hemiptera: Aphididae: Eriosomatinae) on *Rhus* hosts (Anacardiaceae). *Molecular Phylogenetics and Evolution*, 69(3), 1146-1158. <https://doi.org/10.1016/j.ympev.2013.08.003>

Rey, O., Danchin, E., Mirouze, M., Loot, C., & Blanchet, S. (2016). Adaptation to global change: a transposable element–epigenetics perspective. *Trends in ecology & evolution*, 31(7), 514-526. <https://doi.org/10.1016/j.tree.2016.03.013>

Rinker, D. C., Specian, N. K., Zhao, S., & Gibbons, J. G. (2019). Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. *Proceedings of the National Academy of Sciences*, 116(27), 13446-13451. <https://doi.org/10.1073/pnas.1901093116>

Rispe, C., Legeai, F., Nabity, P. D., Fernández, R., Arora, A. K., Baa-Puyoulet, P., ... & Tagu, D. (2020). The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC biology*, 18(1), 1-25. <https://doi.org/10.1186/s12915-020-00820-5>

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *Cran r*, 538, 113-120. [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://cran-r.c3sl.ufpr.br/web/packages/MASS/MASS.pdf](chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://cran.r.c3sl.ufpr.br/web/packages/MASS/MASS.pdf)

Rocafort, M., Bowen, J. K., Hassing, B., Cox, M. P., McGreal, B., de la Rosa, S., ... & Mesarich, C. H. (2022). The *Venturia inaequalis* effector repertoire is dominated by expanded families with predicted structural similarity, but unrelated sequence, to avirulence proteins from other plant-pathogenic fungi. *BMC biology*, 20(1), 1-24. <https://doi.org/10.1186/s12915-022-01442-9>

Robertson, H. M., Warr, C. G., & Carlson, J. R. (2003). Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 100(suppl_2), 14537-14542. <https://doi.org/10.1073/pnas.233584710>

Robertson, H. M., & Wanner, K. W. (2006). The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome research*, 16(11), 1395-1403. <https://doi.org/10.1101/gr.5057506>

Robertson, H. M., Robertson, E. C., Walden, K. K., Enders, L. S., & Miller, N. J. (2019). The chemoreceptors and odorant binding proteins of the soybean and pea aphids. *Insect biochemistry and molecular biology*, 105, 69-78. <https://doi.org/10.1016/j.ibmb.2019.01.005>

Rodriguez, P. A., Escudero-Martinez, C., & Bos, J. I. (2017). An aphid effector targets trafficking protein VPS52 in a host-specific manner to promote virulence. *Plant physiology*, 173(3), 1892-1903. <https://doi.org/10.1104/pp.16.01458>

Rogers, R. L., Cridland, J. M., Shao, L., Hu, T. T., Andolfatto, P., & Thornton, K. R. (2015). Tandem duplications and the limits of natural selection in *Drosophila yakuba* and *Drosophila simulans*. *PLoS One*, 10(7), e0132184. <https://doi.org/10.1371/journal.pone.0132184>

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., Van de Wouw, A. P., Couloux, A., ... & Howlett, B. J. (2011). Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature communications*, 2(1), 202. <https://doi.org/10.1038/ncomms1189>

Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature methods*, 17(2), 155-158. <https://doi.org/10.1038/s41592-019-0669-3>

Sánchez-Gracia, A., Vieira, F. G., & Rozas, J. (2009). Molecular evolution of the major chemosensory gene families in insects. *Heredity*, 103(3), 208-216. <https://doi.org/10.1038/hdy.2009.55>

Santamaria, M. E., Arnaiz, A., Gonzalez-Melendi, P., Martinez, M., & Diaz, I. (2018). Plant perception and short-term responses to phytophagous insects and mites. *International journal of molecular sciences*, 19(5), 1356. <https://doi.org/10.3390/ijms19051356>

- Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., ... & Ostell, J. (2019). Database resources of the national center for biotechnology information. *Nucleic acids research*, 47(Database issue), D23. <https://doi.org/10.1093/nar/gky1069>
- Scott, K., Brady Jr, R., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C., & Axel, R. (2001). A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell*, 104(5), 661-673. [https://doi.org/10.1016/S0092-8674\(01\)00263-X](https://doi.org/10.1016/S0092-8674(01)00263-X)
- Schrader, L., & Schmitz, J. (2019). The impact of transposable elements in adaptive evolution. *Molecular Ecology*, 28(6), 1537-1549. <https://doi.org/10.1111/mec.14794>
- Seidl, M. F., & Thomma, B. P. (2017). Transposable elements direct the coevolution between plants and microbes. *Trends in Genetics*, 33(11), 842-851. <https://doi.org/10.1016/j.tig.2017.07.003>
- Sheffield, N. C., & Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*, 32(4), 587-589. <https://doi.org/10.1093/bioinformatics/btv612>
- Schoonhoven, L. M., Van Loon, B., van Loon, J. J., & Dicke, M. (2005). *Insect-plant biology*. Oxford University Press on Demand. <https://doi.org/10.1641/B580406>
- Schröder, R., & Hilker, M. (2008). The relevance of background odor in resource location by insects: a behavioral approach. *Bioscience*, 58(4), 308-316. <https://doi.org/10.1641/B580406>
- Seong, K., & Krasileva, K. V. (2023). Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. *Nature Microbiology*, 8(1), 174-187. <https://doi.org/10.1038/s41564-022-01287-6>
- Sheppard, S. K., Guttman, D. S., & Fitzgerald, J. R. (2018). Population genomics of bacterial host adaptation. *Nature Reviews Genetics*, 19(9), 549-565. <https://doi.org/10.1038/s41576-018-0032-z>
- Simola, D. F., Ye, C., Mutti, N. S., Dolezal, K., Bonasio, R., Liebig, J., ... & Berger, S. L. (2013). A chromatin link to caste identity in the carpenter ant *Camponotus floridanus*. *Genome Research*, 23(3), 486-496. <https://doi.org/10.1101/gr.148361.112>
- Simon, J. C., d'Alençon, E., Guy, E., Jacquín-Joly, E., Jaquier, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in functional genomics*, 14(6), 413-423. <https://doi.org/10.1093/bfgp/elv015>
- Singh, K. S., Cordeiro, E. M., Troczka, B. J., Pym, A., Mackisack, J., Mathers, T. C., ... & Bass, C. (2021). Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest *Myzus persicae*. *Communications biology*, 4(1), 847. <https://doi.org/10.1038/s42003-021-02373-x>

- Sironi, M., Cagliani, R., Forni, D., & Clerici, M. (2015). Evolutionary insights into host–pathogen interactions from mammalian sequence data. *Nature Reviews Genetics*, 16(4), 224–236. <https://doi.org/10.1038/nrg3905>
- Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics*, 6(1), 1–11. <https://doi.org/10.1186/1471-2105-6-31>
- Smadja, C., Shi, P., Butlin, R. K., & Robertson, H. M. (2009). Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Molecular biology and evolution*, 26(9), 2073–2086. <https://doi.org/10.1093/molbev/msp116>
- Smadja, C., & Butlin, R. K. (2009). On the scent of speciation: the chemosensory system and its role in premating isolation. *Heredity*, 102(1), 77–97. <https://doi.org/10.1038/hdy.2008.55>
- Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013–2015. Institute for Systems Biology. <http://www.repeatmasker.org>
- Sparks, J. T., Vinyard, B. T., & Dickens, J. C. (2013). Gustatory receptor expression in the labella and tarsi of *Aedes aegypti*. *Insect biochemistry and molecular biology*, 43(12), 1161–1171. <https://doi.org/10.1016/j.ibmb.2013.10.005>
- Sproul, J., Hotaling, S., Heckenhauer, J., Powell, A., Marshall, D., Larracuenta, A. M., ... & Frandsen, P. B. (2023). 600+ insect genomes reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation challenges. *Genome Research*, gr-277387. <https://www.biorxiv.org/content/10.1101/2022.06.02.494618v2.full.pdf>
- Srinivasan, D. G., & Brisson, J. A. (2012). Aphids: a model for polyphenism and epigenetics. *Genetics Research International*, 2012. <https://doi.org/10.1155/2012/431531>
- Stahl, E. A., & Bishop, J. G. (2000). Plant–pathogen arms races at the molecular level. *Current opinion in plant biology*, 3(4), 299–304. [https://doi.org/10.1016/S1369-5266\(00\)00083-2](https://doi.org/10.1016/S1369-5266(00)00083-2)
- Stern, David L. 2008. "Aphids." *Current Biology*: CB 18 (12): R504–5. <https://doi.org/10.1016/j.cub.2008.03.034>
- Stergiopoulos, I., & de Wit, P. J. (2009). Fungal effector proteins. *Annual review of phytopathology*, 47, 233–263. <https://doi.org/10.1146/annurev.phyto.112408.132637>
- Suh, E., Bohbot, J. D., & Zwiebel, L. J. (2014). Peripheral olfactory signaling in insects. *Current Opinion in Insect Science*, 6, 86–92. <https://doi.org/10.1016/j.cois.2014.10.006>
- Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., ... & Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome research*, 24(12), 1963–1976. <http://www.genome.org/cgi/doi/10.1101/gr.168872.113>

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*, 34(suppl_2), W609-W612. <https://doi.org/10.1093/nar/gkl315>

Suzuki, Y., & Nei, M. (2002). Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. *Molecular biology and evolution*, 19(11), 1865-1869. <https://doi.org/10.1093/oxfordjournals.molbev.a004010>

Thissen, D., Steinberg, L., & Wainer, H. (2013). Use of item response theory in the study of group differences in trace lines. In *Test validity* (pp. 147-169). Routledge.

Thompson, J. N. (1994). *The coevolutionary process*. University of Chicago press.

Thorpe, P., Escudero-Martinez, C. M., Cock, P. J., Eves-van den Akker, S., & Bos, J. I. (2018). Shared transcriptional control and disparate gain and loss of aphid parasitism genes. *Genome biology and evolution*, 10(10), 2716-2733. <https://doi.org/10.1093/gbe/evy183>

Torres, D. E., Oggenfuss, U., Croll, D., & Seidl, M. F. (2020). Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews*, 34(3), 136-143. <https://doi.org/10.1016/j.fbr.2020.07.001>

Tsagkogeorga, G., Müller, S., Dessimoz, C., & Rossiter, S. J. (2017). Comparative genomics reveals contraction in olfactory receptor genes in bats. *Scientific reports*, 7(1), 259. <https://doi.org/10.1038/s41598-017-00132-9>

Quan, Q., Hu, X., Pan, B., Zeng, B., Wu, N., Fang, G., ... & Zhan, S. (2019). Draft genome of the cotton aphid *Aphis gossypii*. *Insect Biochemistry and Molecular Biology*, 105, 25-32. <https://doi.org/10.1016/j.ibmb.2018.12.007>

Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with S-PLUS*. Springer Science & Business Media.

Vieira, F. G., Sánchez-Gracia, A., & Rozas, J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: purifying selection and birth-and-death evolution. *Genome biology*, 8(11), 1-16. <https://doi.org/10.1186/gb-2007-8-11-r235>

Vieira, F. G., & Rozas, J. (2011). Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome biology and evolution*, 3, 476-490. <https://doi.org/10.1093/gbe/evr033>

Wagner, A. (1998). The fate of duplicated genes: loss or new function?. *BioEssays*, 20(10), 785-788. [https://doi.org/10.1002/\(SICI\)1521-1878\(199810\)20:10%3C785::AID-BIES2%3E3.0.CO;2-M](https://doi.org/10.1002/(SICI)1521-1878(199810)20:10%3C785::AID-BIES2%3E3.0.CO;2-M)

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., & Yu, J. (2010). KaKs_Calculator 2.0: a toolkit

incorporating gamma-series methods and sliding window strategies. *Genomics, proteomics & bioinformatics*, 8(1), 77-80. [https://doi.org/10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3)

Wang, Y., Li, J., & Paterson, A. H. (2013). MCSanX-transposed: detecting transposed gene duplications based on multiple colinearity scans. *Bioinformatics*, 29(11), 1458-1460. <https://doi.org/10.1093/bioinformatics/btt150>

Wang, W., Dai, H., Zhang, Y., Chandrasekar, R., Luo, L., Hiromasa, Y., ... & Cui, F. (2015). Armet is an effector protein mediating aphid-plant interactions. *The FASEB Journal*, 29(5), 2032-2045. <https://doi.org/10.1096/fj.14-266023>

Wang, S., Bai, B., Gao, B., & Liu, J. (2019). Comparing performance of *Aphis glycines* Matsumura fed on two novel hosts relative to *Glycine max* (L.) Merrill. *Journal of Asia-Pacific Entomology*, 22(3), 975-981. <https://doi.org/10.1016/j.aspen.2019.07.005>

Warren, R. L., Coombe, L., Mohamadi, H., Zhang, J., Jaquish, B., Isabel, N., ... & Birol, I. (2019). ntEdit: scalable genome sequence polishing. *Bioinformatics*, 35(21), 4430-4432. <https://doi.org/10.1093/bioinformatics/btz400>

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., ... & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution*, 35(3), 543-548. <https://doi.org/10.1093/molbev/msx319>

Weingartner, E., Wahlberg, N., & Nylin, S. (2006). Dynamics of host plant use and species diversity in Polygonia butterflies (Nymphalidae). *Journal of evolutionary biology*, 19(2), 483-491. <https://doi.org/10.1111/j.1420-9101.2005.01009.x>

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome research*, 27(5), 757-767. <http://www.genome.org/cgi/doi/10.1101/gr.214874.116>

White, P. S., Arslan, D., Kim, D., Penley, M., & Morran, L. (2021). Host genetic drift and adaptation in the evolution and maintenance of parasite resistance. *Journal of Evolutionary Biology*, 34(5), 845-851. <https://doi.org/10.1111/jeb.13785>

Whiteman, N. K., & Pierce, N. E. (2008). Delicious poison: genetics of *Drosophila* host plant preference. *Trends in Ecology & Evolution*, 23(9), 473-478. <https://doi.org/10.1016/j.tree.2008.05.010>

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., ... & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12), 973-982. <https://doi.org/10.1038/nrg2165>

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics. Version, 2(1)*, 1-189.

<https://github.com/hadley/ggplot2>

Will, T., & Vilcinskas, A. (2015). The structural sheath protein of aphids is required for phloem feeding. *Insect biochemistry and molecular biology*, 57, 34-40. <https://doi.org/10.1016/j.ibmb.2014.12.005>

Williams, I. S., & Dixon, A. F. (2007). Life cycles and polymorphism. *Aphids as crop pests*, 69-85.

Wyka, S. A., Mondo, S. J., Liu, M., Dettman, J., Nalam, V., & Broders, K. D. (2021). Whole-genome comparisons of ergot fungi reveals the divergence and evolution of species within the genus *Claviceps* are the result of varying mechanisms driving genome evolution and host range expansion. *Genome biology and evolution*, 13(2), evaa267. <https://doi.org/10.1093/gbe/evaa267>

Xiao, J. H., Yue, Z., Jia, L. Y., Yang, X. H., Niu, L. H., Wang, Z., ... & Huang, D. W. (2013). Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome biology*, 14, 1-18. <https://doi.org/10.1186/gb-2013-14-12-r141>

Xu, T. T., Jiang, L. Y., Chen, J., & Qiao, G. X. (2020). Host plants influence the symbiont diversity of Eriosomatinae (Hemiptera: Aphididae). *Insects*, 11(4), 217. <https://doi.org/10.3390/insects11040217>

Yang, Z. (1998). On the best evolutionary rate for phylogenetic analysis. *Systematic Biology*, 47(1), 125-133. <https://doi.org/10.1080/106351598261067>

Yang, Z., & Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of molecular evolution*, 46, 409-418. <https://doi.org/10.1007/PL00006320>

Yang, Z., & Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular biology and evolution*, 19(6), 908-917. <https://doi.org/10.1093/oxfordjournals.molbev.a004148>

Yang, Z., Nielsen, R., Goldman, N., & Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, 155(1), 431-449. <https://doi.org/10.1093/genetics/155.1.431>

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular biology and evolution*, 22(4), 1107-1118. <https://doi.org/10.1093/molbev/msi097>

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, 24(8), 1586-1591. <https://doi.org/10.1093/molbev/msm088>

Yang, S., Cao, D., Wang, G., & Liu, Y. (2017). Identification of genes involved in chemoreception in *Plutella xylostella* by antennal transcriptome analysis. *Scientific Reports*, 7(1), 1-16. <https://doi.org/10.1038/s41598-017-11646-7>

Yates, A. D., & Michel, A. (2018). Mechanisms of aphid adaptation to host plant resistance. *Current opinion in insect science*, 26, 41-49. <https://doi.org/10.1016/j.cois.2018.01.003>

Yeo, S., Coombe, L., Warren, R. L., Chu, J., & Birol, I. (2018). ARCS: scaffolding genome drafts with linked reads. *Bioinformatics*, 34(5), 725-731. <https://doi.org/10.1093/bioinformatics/btx675>

Zhan, J., Mundt, C. C., Hoffer, M. E., & McDonald, B. A. (2002). Local adaptation and effect of host genotype on the rate of pathogen evolution: an experimental test in a plant pathosystem. *Journal of Evolutionary Biology*, 15(4), 634-647. <https://doi.org/10.1046/j.1420-9101.2002.00428.x>

Zhang, J. (2004). Frequent false detection of positive selection by the likelihood method with branch-site models. *Molecular biology and evolution*, 21(7), 1332-1339. <https://doi.org/10.1093/molbev/msh117>

Zhang, Y. N., Zhu, X. Y., Ma, J. F., Dong, Z. P., Xu, J. W., Kang, K., & Zhang, L. W. (2017). Molecular identification and expression patterns of odorant binding protein and chemosensory protein genes in *Athetis lepigone* (Lepidoptera: Noctuidae). *PeerJ*, 5, e3157. <http://doi.org/10.7717/peerj.3157>

Zhang, S., Gao, X., Wang, L., Jiang, W., Su, H., Jing, T., ... & Yang, Y. (2022). Chromosome-level genome assemblies of two cotton-melon aphid *Aphis gossypii* biotypes unveil mechanisms of host adaptation. *Molecular Ecology Resources*, 22(3), 1120-1134. <https://doi.org/10.1111/1755-0998.13521>

Zhang, S., Li, C., Si, J., Han, Z., & Chen, D. (2022). Action Mechanisms of Effectors in Plant-Pathogen Interaction. *International Journal of Molecular Sciences*, 23(12), 6758. <https://doi.org/10.3390/ijms23126758>

Zidi, M., Klai, K., Confais, J., Chénais, B., Caruso, A., Denis, F., ... & Casse, N. (2022). Genome-Wide Screening of Transposable Elements in the Whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae), Revealed Insertions with Potential Insecticide Resistance Implications. *Insects*, 13(5), 396. <https://doi.org/10.3390/insects13050396>

Zytynska, S. E., Frantz, L., Hurst, B., Johnson, A., Preziosi, R. F., & Rowntree, J. K. (2014). Host-plant genotypic diversity and community genetic interactions mediate aphid spatial distribution. *Ecology and evolution*, 4(2), 121-131. <https://doi.org/10.1002/ece3.916>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

4.2 TEXT S1. DNA EXTRACTION AND *DYSAPHIS PLANTAGINEA* GENOME SEQUENCING AND ASSEMBLY

We used a clone (16-0042-0001) from a colony of *Dysaphis plantaginea* Passerini that was collected on *Malus domestica* Borkh in 2016 in Torchefelon (France). We reared on *Plantago lanceolata* L. under control conditions (16h light, 8h dark at 20°C and 75% of humidity) at the *Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail* (ANSES, Lyon, France). We extracted high molecular weight DNA from clone 16_042 at the *Centre National de Ressources Génomiques Végétales* (CNRGV Toulouse, France). We sequenced it using Chromium 10X Genomics technology (Genopole Toulouse, Toulouse, France).

We assembled the 10X reads using Supernova™ (Weisenfeld et al., 2017). We defined two assemblies as 56X and 100X due to the number of fragments involved during the process. We build the scaffolds using the ARCS (Assembly Round-up by Chromium Scaffolding) algorithm (Yeo et al., 2018). We assessed the quality of the *D. plantaginea* assembly version 1 using Benchmarking Universal Single-Copy Orthologs BUSCO (Waterhouse et al., 2018). We confronted the two assemblies (56X and 100X) with BWA (Li and Durbin, 2009) and Samtools (Li et al., 2009) (Table S1). We utilized this first version of the *D. plantaginea* genome assembly to extract the complete *Buchnera aphidicola* Munson and mitochondrial genomes. We reconstructed the *B. aphidicola* genome with two methods. First, we identified the scaffolds by blasting the assemblages against the bacterium *B. aphidicola* of *Myzus persicae* Sulzer reference library. Second, we filled the gaps using the raw data, and the scaffolds aligned on the *B. aphidicola* genome of *M. persicae*.

27 We improved the *Dysaphis plantaginea* genome assembly v1 using long
28 reads obtained from Pacific Biosciences (PacBio) and Oxford Nanopore Technolo-
29 gies (ONT). We used the wtdbg2 v2.5 (Ruan and Li, 2020) on long reads and ntEdit
30 (Warren et al., 2019) on the 10X Illumina sequences to search for inconsistencies
31 (e.g., local misassemblies). We constructed the scaffolds with Tigmint (Jackman et
32 al., 2018) and ARCS (Yeo et al., 2018) with the 10X data. Finally, we re-assembled
33 the genome using FLYE (Kolmogorov et al., 2019), filtering out the reads from *B.*
34 *aphidicola*. Afterward, we applied the software and procedure of the second ver-
35 sion of the *D. plantaginea* genome assembly to obtain the version 3 of the genome
36 (Table S1). We constructed the optical mapping using MapSolver (OpGen®) to un-
37 derstand the genomic structure and structural variation. We performed the gene
38 annotation using Maker2 (Holt and Yandell, 2011) and the TE annotation with
39 REPET (see details in the material and methods Repetitive elements annotation
40 section; Flutre et al., 2011).

41

42 We detected and removed the mitochondrial genome sequence based on
43 the similarity with *M. persicae* during genome assembly. The initial mitochondrial
44 scaffold was 18.8 Kbp in length. We performed the gene prediction on this scaffold
45 with MITOS (Bernt et al., 2013) and ARWEN v1.2 (Laslett and Canback, 2008).

46

47 *Dysaphis plantaginea* genome assembly and gene annotation

48 The *D. plantaginea* genome v3 represented the twelfth released genome in
49 the Aphidinae subfamily on the AphidBase-BIPAA platform. The *D. plantaginea* ge-
50 nome assembly version 3 summed a size of 485.98 Mb (The different version of
51 the *D. plantaginea* new assembly were presented in Table S1a and the assembly
52 statistic of the *D. plantaginea* V3 were presented in Table S1b). The genome as-
53 sembly comprised 2,366 scaffolds with a mean scaffold size of 205,401 bp and an

54 N50 value of 17,731,133 bp. BUSCO analysis based on hemipteran conserved
55 genes indicated the presence of 99.4% complete genes. We detected a genomic
56 GC content of 30.14% and annotated 53,848 genes and 18,665 proteins.

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80 REFERENCES

81 Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011). Considering transpo-
82 sable element diversification in de novo annotation approaches. *PloS one*, 6(1),
83 e16526. <https://doi.org/10.1371/journal.pone.0016526>

84

85 Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-data-
86 base management tool for second-generation genome projects. *BMC bioinformat-*
87 *ics*, 12(1), 1-14. <https://doi.org/10.1186/1471-2105-12-491>

88

89 Jackman, S. D., Coombe, L., Chu, J., Warren, R. L., Vandervalk, B. P., Yeo, S., ... &
90 Birol, I. (2018). Tigmint: correcting assembly errors using linked reads from large
91 molecules. *BMC bioinformatics*, 19(1), 1-10. [https://doi.org/10.1186/s12859-018-](https://doi.org/10.1186/s12859-018-2425-6)
92 [2425-6](https://doi.org/10.1186/s12859-018-2425-6)

93

94 Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-
95 prone reads using repeat graphs. *Nature Biotechnology*, 37(5), 540-546.
96 <https://doi.org/10.1038/s41587-019-0072-8>

97

98 Laslett, D., & Canbäck, B. (2008). ARWEN: a program to detect tRNA genes in met-
99 azoan mitochondrial nucleotide sequences. *Bioinformatics*, 24(2), 172-175.
100 <https://doi.org/10.1093/bioinformatics/btm573>

101

102 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-
103 Wheeler transform. *bioinformatics*, 25(14), 1754-1760. [https://doi.org/10.1093/bi-](https://doi.org/10.1093/bioinformatics/btp324)
104 [oinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324)

105

106 Li, Y., Park, H., Smith, T. E., & Moran, N. A. (2019). Gene family evolution in the pea
107 aphid based on chromosome-level genome assembly. *Molecular biology and evo-*
108 *lution*, 36(10), 2143-2156. <https://doi.org/10.1093/molbev/msz138>

109

110 Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsche, G., ... &
111 Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome
112 annotation. *Molecular phylogenetics and evolution*, 69(2), 313-319.
113 <https://doi.org/10.1016/j.ympev.2012.08.023>

114

115 Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature*
116 *methods*, 17(2), 155-158. <https://doi.org/10.1038/s41592-019-0669-3>

117

118 Warren, R. L., Coombe, L., Mohamadi, H., Zhang, J., Jaquish, B., Isabel, N., ... & Birol,
119 I. (2019). ntEdit: scalable genome sequence polishing. *Bioinformatics*, 35(21), 4430-
120 4432. <https://doi.org/10.1093/bioinformatics/btz400>
121
122 Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov,
123 G., ... & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to
124 gene prediction and phylogenomics. *Molecular biology and evolution*, 35(3), 543-
125 548. <https://doi.org/10.1093/molbev/msx319>
126
127 Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct
128 determination of diploid genome sequences. *Genome research*, 27(5), 757-767.
129 <http://www.genome.org/cgi/doi/10.1101/gr.214874.116>
130
131 Yeo, S., Coombe, L., Warren, R. L., Chu, J., & Birol, I. (2018). ARCS: scaffolding ge-
132 nome drafts with linked reads. *Bioinformatics*, 34(5), 725-731.
133 <https://doi.org/10.1093/bioinformatics/btx675>
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157

158 4.3 SUPPLEMENTARY MATERIAL

159 **Text S1.** DNA extraction and *Dysaphis plantaginea* genome sequencing and assembly

160 **Table S1.** Main statistics of the *Dysaphis plantaginea* assembly versions 1, 2 and 3

161

162 **Table S2.** Statistics for the *Dysaphis plantaginea* genome assembly version 3, and
163 transposable elements content.

164

165 **Table S3.** Details of the 13 genomes, genome sizes, assembly quality used in this study,
166 and associated chemosensory gene and transposable contents were estimated.

167

168 **Table S4.** Summary of the orthogroups in general and per species detected by
169 OrthoFinder V (Emms and Kelly, 2019).

170

171 **Table S5.** Statistical test of the differences in genes numbers among species and gene
172 family (OR and GR) annotated in the 13 aphid species. Gene count was fitted to a
173 Negative Binomial generalized linear model (glm.nb function).

174

175 **Table S6.** Information about the aphid host preferences and life cycle was used in this
176 study.

177

178 **Table S7.** Likelihood-ratio test of models with different evolutionary rates of the 12
179 aphid and one aphid-like species depending on the taxonomic ranks (genus,
180 subfamilies, families, species) or host preferences (see details in Table S5) or with a
181 single evolutionary rate along the phylogeny inferred with CAFE (Han et al., 2013).

182

183 **Table S8.** Enrichment analysis of expanded genes.

184

185 **Table S9.** Enrichment analysis of contracted genes.

186

187 **Table S10.** Gene detected by Syntenet (Almeida et al., 2023) clustering with
188 orthogroup (OrthoFinder, Emms and Kelly, 2019) and chemosensory gene.

189

190 **Table S11.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors -
191 OR) branch model summary statistics inferred from PAML v4 (Yang, 2007).

192

193 **Table S12.** Chemosensory genes (gustatory receptors GR and olfactory receptors OR)
194 were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection
195 based on the branch model analysis PAML v4 (Yang, 2007).

196

197 **Table S13.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors -

198 OR) site model summary statistics inferred from PAML v4 (Yang, 2007).
199
200 **Table S14.** Chemosensory genes (gustatory genes-GR and olfactory genes-OR)
201 likelihood-rate test summary of the different selection test models.
202
203 **Table S15.** Chemosensory genes (gustatory receptors GR and olfactory receptors OR)
204 were suggested to evolve under positive and purifying selection based on the site
205 model analysis.
206
207 **Table S16.** Selection test summary based on the branch model using single-copy
208 orthologous groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and
209 neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang, 2007).
210
211 **Table S17.** Single-copy orthologous groups (OG) were suggested under positive
212 ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model
213 analysis PAML v4 (Yang, 2007).
214
215 **Table S18.** Single-copy orthologous groups (OG) selection test summary based on the
216 branch model using single-copy orthologous groups were suggested under positive
217 ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model
218 analysis PAML v4 (Yang, 2007).
219
220 **Table S19.** Single-copy orthologous groups (OG) likelihood-rate test summary of the
221 different selection test models.
222
223 **Table S20.** Single-copy orthologous groups (OG) suggested sites to evolve under
224 positive and purifying selection based on the site model analysis.
225
226 **Table S21.** Statistics of the genome assembly using for the REPET pipeline of the 13
227 aphid and aphid-like species.
228
229 **Table S22. Statistics** of the transposable elements consensus library from the REPET
230 pipeline of the 13 aphid and aphid-like species.
231
232 **Table S23.** General linear model (GLM) summary of different models to test the
233 distribution of the selection pressure among the different chemosensory genes and
234 the single-copy orthologous gene groups (OG).
235
236 **Table S24.** Locus Overlap Analysis (LOLA) transposable elements (TEs) enrichment
237 analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes in a 10Kb
238 length. Excel file:
239

240 **Table S25.** Locus Overlap Analysis (LOLA) transposable elements (TEs) depletion
241 analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes in a 10Kb
242 length.

243

244 **Table S26.** Locus Overlap Analysis (LOLA) transposable elements (TEs) enrichment and
245 depletion analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes in
246 a 2Kb length.

247

248

249 **Figure S1.** Gene orthology among aphid and one aphid-like species. Bars represent
250 the gene number of different types of orthologues: multi-copy orthologous genes in
251 all species, species-specific (unique genes to specific species); unassigned (genes
252 cannot be assigned to any orthogroups).

253

254 **Figure S2.** Chemosensory genes manual annotation workflow.

255

256 **Figure S3.** Number of olfactory and gustatory genes detected for the 12 aphid species
257 and one aphid-like species, and single copy orthologous groups used as references for
258 the genome background.

259

260 **Figure S4.** Number of olfactory and gustatory genes detected for the 12 aphid species
261 and one aphid-like species predicted with Model 1 (Table S5).

262

263 **Figure S5.** Enrichment analysis of genes underwent expansion during the aphid and
264 aphid-like species.

265

266 **Figure S6.** Enrichment analysis of genes underwent contraction during the aphid and
267 aphid-like species.

268

269 **Figure S7.** Twelve groups of the candidate gustatory receptors (GR) genes were
270 detected using a phylogenetic tree.

271

272 **Figure S8.** Twenty-four groups of the candidate olfactory receptors (OR) genes were
273 detected using a phylogenetic tree.

274

275 **Figure S9.** Chemosensory genes (OR and GR) and single-copy orthologous genes
276 dNdS distribution suggest purifying selection.

277

278 **Figure S10.** Correlation between dN/dS ratio and synonymous divergence (dS) among
279 Gustatory Receptors (GR), Olfactory receptors (OR), and single copy orthologous
280 groups (OG) for the 13 species.

281

282 **Figure S11.** Changes in effective population size (N_e) estimates from pairwise
283 sequentially Markovian coalescent (PSMC; [Li and Durbin, 2011](#)) with unmasked and
284 masked genomes for repeat regions for each aphid species.

285

286 **Figure S12.** TE-copy identity number among the 12 aphid species using an upset plot
287 (UpSetR R package, Conway et al., 20

288 **Table S1.** Main statistics of *the Dysaphis plantaginea* assemblies, versions 1, 2, and 3.

289

<i>Dysaphis plantaginea</i> version	Assembler	Number scaffolds	Scaffolds (Mb)	Longest scaffold (Mb)	N50 (Mb)	L50	BUSCO genes	Complete and single-copy BUSCOs (S)	Complete and duplicated BUSCOs (D)	Fragments BUSCOs (F)	Missing BUSCOs (M)
1	Supernova 100X	24,694	446.4	10.7	2.5	48	94.1%	22,655	NA	337	1,083
1	Supernova 56X	24,075	389.2	15.2	2.1	41	92.9%	22,366		554	1,156
2	Wtdbg	3,458	411.7	13.8	2.5	40	~75%	1,188	48	14	117
3	FLYE	2,366	486	88.6	17.7	7	~96%	1,251	63	10	43

290

291 **N50:** the sequence length of the shortest contig at 50% of the total assembly length; **L50:** count of the smallest number of contigs whose length sum makes up half of genome size;
292 **BUSCO:** Benchmarking Universal Single-Copy Orthologs.

293

294

295

296

297

298

299

300

301 **Table S2. Statistics for the *Dysaphis plantaginea* genome assembly version 3, assembly completeness assessed by BUSCO,**
 302 **and transposable elements content**

Genome size		485 Mb			
Assembly information		Contig information		BUSCO analysis	
Total assembly size	485,979,193	Num contigs	3,398	Completed BUSCOs (C)	2,495 (99.4%)
Num scaffolds	2,366	Total contig size	441,227,050	Completed and single-copy BUSCOs (S)	2,424 (96.6%)
Longest scaffold	88,606,889	Longest contig	5,044,119	Completed and duplicated BUSCOs (D)	71 (2.8%)
Shorter scaffold	13	Shorter contigs	12	Fragmented BUSCOs (F)	4 (0.2%)
Mean scaffold size	205,401	Mean contig size	12,9849	Missing BUSCOs (M)	11 (0.4%)
N50 scaffold (bp)	17,731,133	N50 contig (bp)	539,163		
L50 scaffold (bp)	7	L50 contig (bp)	186		
GC (%)	27.37	GC (%)	30.14		
				Transposable elements content	
				TEs in genome (Mb)	255
				TEs in genome (%)	63.78

303

304

305

306

307 **Table S3.** Details of the 13 genomes, genome sizes, assembly quality used in this study, and associated chemosensory genes and
 308 transposable contents estimated with REPET (Quesneville et al., 2005; Flutre et al., 2011).

309

Aphid species information					
Species	Common name	Acronyms	Publication	Raw data coverage for assembly	Genome Size (Mb)
<i>Aphis glycines</i> (clone BT1)	Soybean aphid	AGLY	Mathers, 2020	5X	308
<i>Aphis gossypii</i>	Cotton aphid	AGOS	Quan et al., 2019	-	294
<i>Acyrtosiphon pisum</i> (clone LSR1)	Pea aphid	APIS	The International Aphid Genomics Consortium, 2010	39X	541
<i>Diuraphis noxia</i>	Russian wheat aphid	DNOX	Nicholson et al., 2015	53X	393
<i>Dysaphis plantaginea</i>	Rosy apple aphid	DPLA	ECLECTIC-GQE-INRAE-CNRS	19X	485
<i>Daktulosphaira vitifoliae</i>	Grape phylloxera	DVIT	Rispe et al., 2020	26X	283
<i>Eriosoma lanigerum</i>	Wolly apple aphid	ELAN	Biello et al., 2021	309X	335
<i>Myzus cerasi</i>	Black cherry aphid	MCER	Thorpe et al., 2018	82X	406
<i>Myzus persicae</i> (clone O)	Green peach aphid	MPER	Mathers et al., 2021	131X	391
<i>Pentalonia nigronervosa</i>	Bannana aphid	PNIG	Mathers et al., 2020	100X	375
<i>Rhopalosiphum maidis</i>	Maize aphid	RMAI	Chen et al., 2019	21X	326
<i>Rhopalosiphum padi</i>	Bird-cherry oat aphid	RPAD	Thorpe et al., 2018	761X	322
<i>Sitobion miscanthi</i>	Indian grain aphid	SMIS	Jiang et al., 2019	172X	398

310

311

312

313 **Table S3 continuation.** Details of the 13 genomes, genome sizes, assembly quality used in this study, and associated chemosensory
 314 genes and transposable contents estimated with REPET (Quesneville et al., 2005; Flutre et al., 2011).

315

Scaffold assembly information								
Acronyms	Total assembly size (bp)	Num scaffolds	Longest scaffold	Shorter scaffold	Mean scaffold size	N50 scaffold (bp)	L50 scaffold (bp)	scaffold GC (%)
AGLY	308,065,118	3,224	23,246,776	60	95,554	5,864,887	15	26.41
AGOS	294,278,726	4,718	5,574,629	889	2,91	437,96	195	25.78
APIS	541,137,574	21,92	170,740,645	200	24,687	132,544,852	2	29.63
DNOX	395,073,589	5,637	2,142,037	928	70,086	397,774	281	21.78
DPLA	485,979,193	2,366	88,606,889	13	205,401	17,731,133	7	27.37
DVIT	282,671,353	10,492	2,080,308	141	26,942	341,59	236	26.63
ELAN	334,868,377	7,146	71,231,741	1	46,861	62,861,676	3	25.35
MCER	405,711,039	49,286	265,361	1,001	8,232	23,273	4472	29.86
MPER	390,790,610	360	105,178,091	5	1,097,619	69,480,500	3	30.21
PNIG	375,348,459	18,348	631,822	1	20,457	103,994	1038	28.64
RMAI	326,023,155	220	94,224,415	1,096	1,481,923	93,298,903	2	27.69
RPAD	321,589,008	2,172	4,088,110	1,131	148,061	652,723	140	27.81
SMIS	397,852,643	653	101,470,385	2,682	609,269	36,263,045	4	30.06

316

317

318

319

320 **Table S3 continuation.** Details of the 13 genomes, genome sizes, assembly quality used in this study, and associated chemosensory
 321 genes and transposable contents estimated with REPET (Quesneville et al., 2005; Flutre et al., 2011).

Acronyms	Contigs assembly information							
	Number of contigs	Total size of contigs	Longest contigs	Shortest contigs	Mean contig size	N50 contig (bp)	L50 contig (bp)	contig GC (%)
AGLY	3,587	302,879,913	7,698,601	60	84,438	1,958,040	47	27.27
AGOS	12,144	278,291,622	709,413	415	22,916	77,914	960	27.26
APIS	68,185	499,739,085	424,12	1	73,29	25,858	4,765	29.65
DNOX	50,723	296,543,584	167,15	60	5,846	13,309	5,772	29.01
DPLA	3,398	441,227,050	5,044,119	12	12,9849	539,163	186	30.14
DVIT	17,162	276,429,103	718,286	83	16,107	74,75	1,012	27.23
ELAN	9,919	327,134,977	2,875,571	3	32,981	60,282	163	25.95
MCER	51,353	405,518,793	209,856	1,001	7,897	19,719	5,425	29.88
MPER	914	394,755,899	18,090,249	2,301	43,1899	4,169,528	23	30.23
PNIG	20,87	375,096,534	400,055	1	3,822	64,061	1,681	28.66
RMAI	689	325,976,255	42,509,236	1,096	45,424	9,046,396	10	27.69
RPAD	2,172	321,589,008	4,088,110	1,131	14,8061	652,723	140	27.81
SMIS	1,145	397,803,443	10,203,765	2,682	34,7427	1,613,763	58	30.06

322

323

324

325

326

327 **Table S3 continuation.** Details of the 13 genomes, genome sizes, assembly quality used in this study, and associated chemosensory
 328 genes and transposable contents estimated with REPET (Quesneville et al., 2005; Flutre et al., 2011).

329

330

331

332

333

334

335

336

337

338

Acronyms	Number of BUSCOs	BUSCO statistics										Total BUSCOs groups searched
		Completed BUSCOs (C)	%C	Completed and single-copy BUSCOs (S)	%S	Completed and Duplicated BUSCOs (D)	%D	Fragmented BUSCOs (F)	%F	Missing BUSCOs (M)	%M	
AGLY	2510	2427	96.7	2404	95.8	23	0.9	0	0.0	83	3.3	2510
AGOS	2510	2442	97.3	2371	94.5	71	2.8	14	0.6	54	2.1	2510
APIS	2510	2486	99.0	2443	97.3	43	1.7	4	0.2	20	0.8	2510
DNOX	2510	2432	96.9	2413	96.1	19	0.8	38	1.5	40	1.6	2510
DPLA	2510	2495	99.4	2424	96.6	71	2.8	4	0.2	11	0.4	2510
DVIT	2510	2480	98.8	2428	96.7	52	2.1	7	0.3	23	0.9	2510
ELAN	2510	2484	99.0	2484	97.7	32	1.3	9	0.4	17	0.6	2510
MCER	2510	2472	96.8	2430	96.8	42	1.7	21	0.8	17	0.7	2510
MPER	2510	2484	99.0	2474	98.6	10	0.4	2	0.1	24	0.9	2510
PNIG	2510	2491	99.3	2477	98.7	14	0.6	7	0.3	12	0.4	2510
RMAI	2510	2489	99.1	2463	98.1	26	1.0	5	0.2	16	0.7	2510
RPAD	2510	2487	99.1	2462	98.1	25	1.0	7	0.3	16	0.6	2510
SMIS	2510	2402	95.7	2313	92.2	89	3.5	7	0.3	101	4.0	2510

339

340

341

342

343 **Table S3 continuation.** Details of the 13 genomes, genome sizes, assembly quality used in this study, and associated chemosensory
 344 genes and transposable contents estimated with REPET (Quesneville et al., 2005; Flutre et al., 2011).

	Chemosensory Genes			Transposable elements (TE)			
	Acronyms	GR	OR	TEs in genome (Mb)	TE (%)	TE Class I (%)	TE Class II(%)
345							
346	AGLY	58	47	170	55.98	39.31	16.67
	AGOS	49	55	57	43.07	18.98	24.09
347	APIS	71	70	150	46.10	21.40	24.70
	DNOX	26	30	50	52.36	26.46	28.90
348	DPLA	40	35	255	53.42	12.94	40.48
	DVIT	18	64	200	34.82	20.31	14.51
349	ELAN	23	43	225	41.35	15.79	25.56
	MCER	40	65	89	45.25	14.83	30.42
350	MPER	56	48	192	53.79	20.78	33.01
	PNIG	37	37	94	64.27	17.95	46.32
351	RMAI	47	52	160	43.24	22.23	21.01
	RPAD	64	61	148	44.66	14.98	29.68
352	SMIS	51	61	184	43.46	19.38	24.58
353							

354

355

356

357

358

359 **Table S4.** Summary of the orthogroups in general and per species detected by OrthoFinder V (Emms and Kelly, 2019).

360

361 **Summary of the orthologous groups detected**

361	Number of species	13
362	Number of genes	380653
362	Number of genes in orthogroups	334936
363	Number of unassigned genes	45717
363	Percentage of genes in orthogroups	88
364	Percentage of unassigned genes	12
364	Number of orthogroups	28672
365	Number of species-specific orthogroups	5713
365	Number of genes in species-specific orthogroups	28227
366	Percentage of genes in species-specific orthogroups	7,4
366	Mean orthogroup size	11,7
367	Median orthogroup size	6
367	G50 (assigned genes)	18
368	G50 (all genes)	16
368	O50 (assigned genes)	4405
369	O50 (all genes)	5741
370	Number of orthogroups with all species present	5868
370	Number of single-copy orthogroups	891

371

372

373

374 **Table S4 continuation.** Summary of the orthogroups in general and per species detected by OrthoFinder V (Emms and Kelly, 2019).

375

	Orthologous groups statistics per species												
	AGLY	AGOS	APIS	DNOX	DPLA	DVIT	ELAN	MCER	MPER	PNIG	RMAI	RPAD	SMIS
Number of genes	42244	18504	28016	17484	53923	25794	28310	28793	31593	29706	19498	26551	30237
Number of genes in orthogroups	28489	18236	27416	17169	46719	23400	23108	24686	29943	27604	19376	22573	26217
Number of unassigned genes	13755	268	600	315	7204	2394	5202	4107	1650	2102	122	3978	4020
Percentage of genes in orthogroups	67,4	98,6	97,9	98,2	86,6	90,7	81,6	85,7	94,8	92,9	99,4	85	86,7
Percentage of unassigned genes	32,6	1,4	2,1	1,8	13,4	9,3	18,4	14,3	5,2	7,1	0,6	15	13,3
Number of orthogroups containing species	15567	10414	11495	10133	18085	10884	12308	14784	13863	13778	10036	13753	15142
Percentage of orthogroups containing species	54,3	36,3	40,1	35,3	63,1	38	42,9	51,6	48,4	48,1	35	48	52,8
Number of species-specific orthogroups	664	43	270	32	1550	719	660	186	288	540	35	280	446
Number of genes in species-specific orthogroups	4975	106	1153	71	5719	6067	3626	506	1272	1855	166	1357	1354
Percentage of genes in species-specific orthogroups	11,8	0,6	4,1	0,4	10,6	23,5	12,8	1,8	4	6,2	0,9	5,1	4,5

376

377

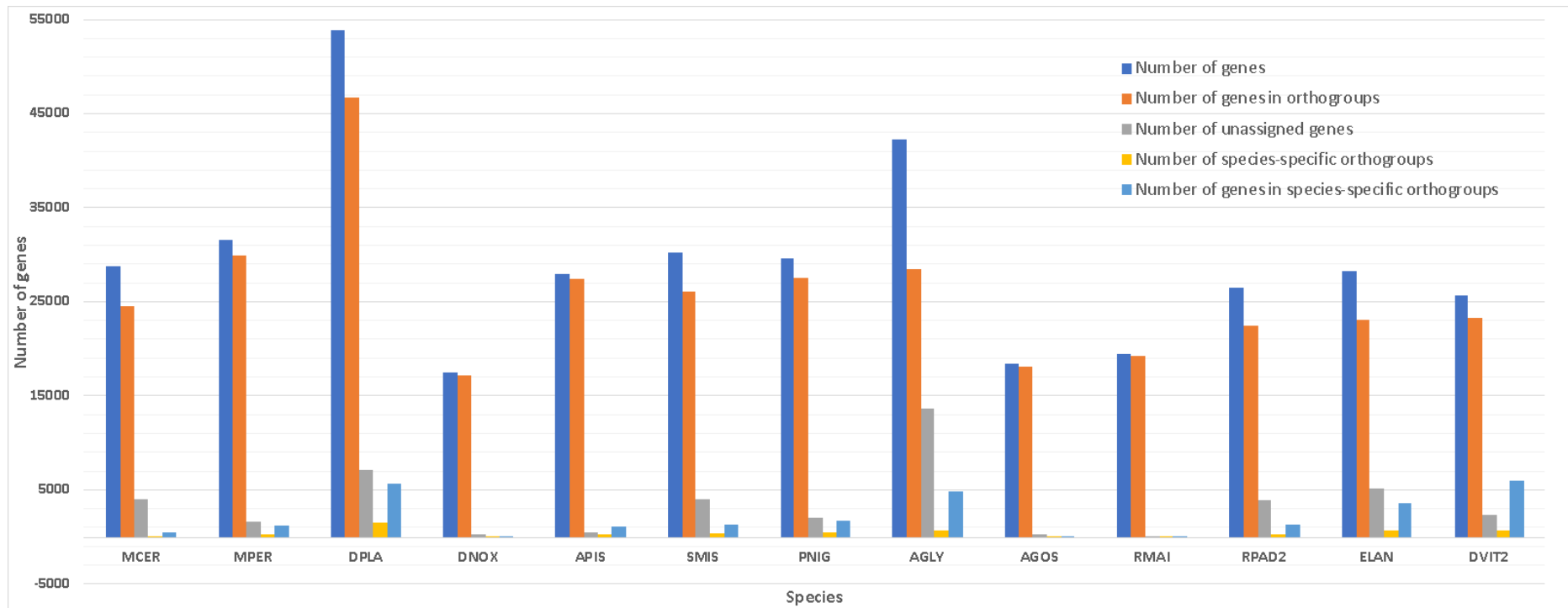
378

379

380

381

382

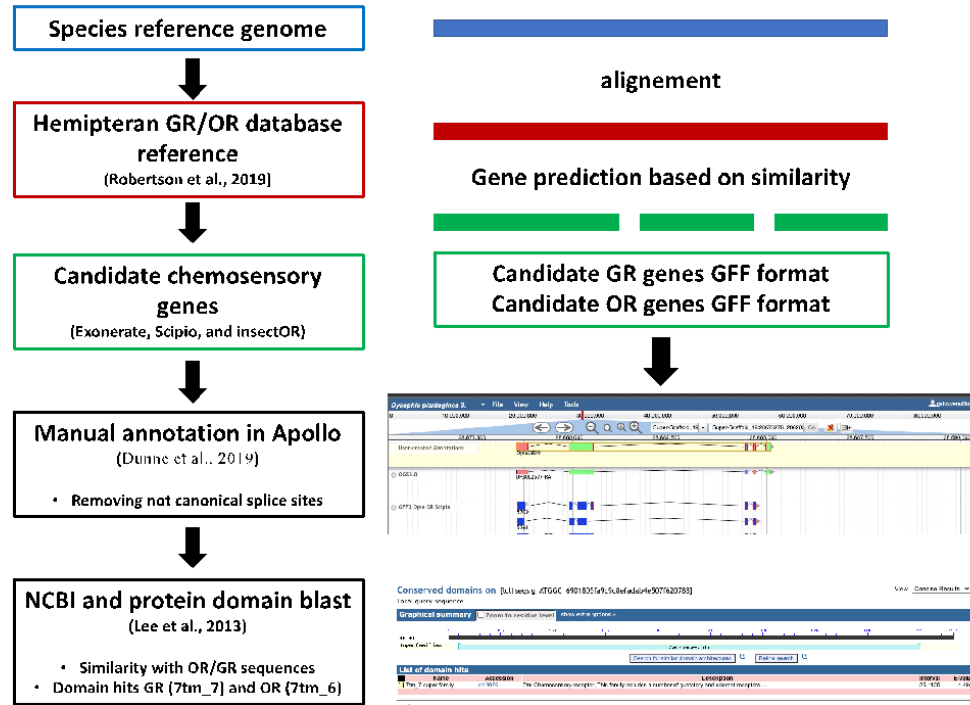


384

385 **Figure S1.** Gene orthology among aphid and one aphid-like species. Bars represent the gene number of different types of
 386 orthologues: multi-copy orthologous genes in all species, species-specific (unique genes to specific species); unassigned (genes
 387 cannot be assigned to any orthogroups). MCER: *Myzus cerasi*; MPER: *Myzus persicae*; DPLA: *Dysaphis plantaginea*; DNOX: *Diuraphis*
 388 *noxia*; APIS: *Acyrtosiphon pisum*; SMIS: *Sitobion miscanthi*; PNIG: *Pentalonia nigronervosa*; AGLY: *Aphis glycines*; AGOS: *Aphis gossypii*;
 389 RMAI: *Rhopalosiphum maidis*; RPAD: *Rhopalosiphum padi*; ELAN: *Eriosoma lanigerum*; DVIT: *Daktulosphaira vitifoliae*.

390

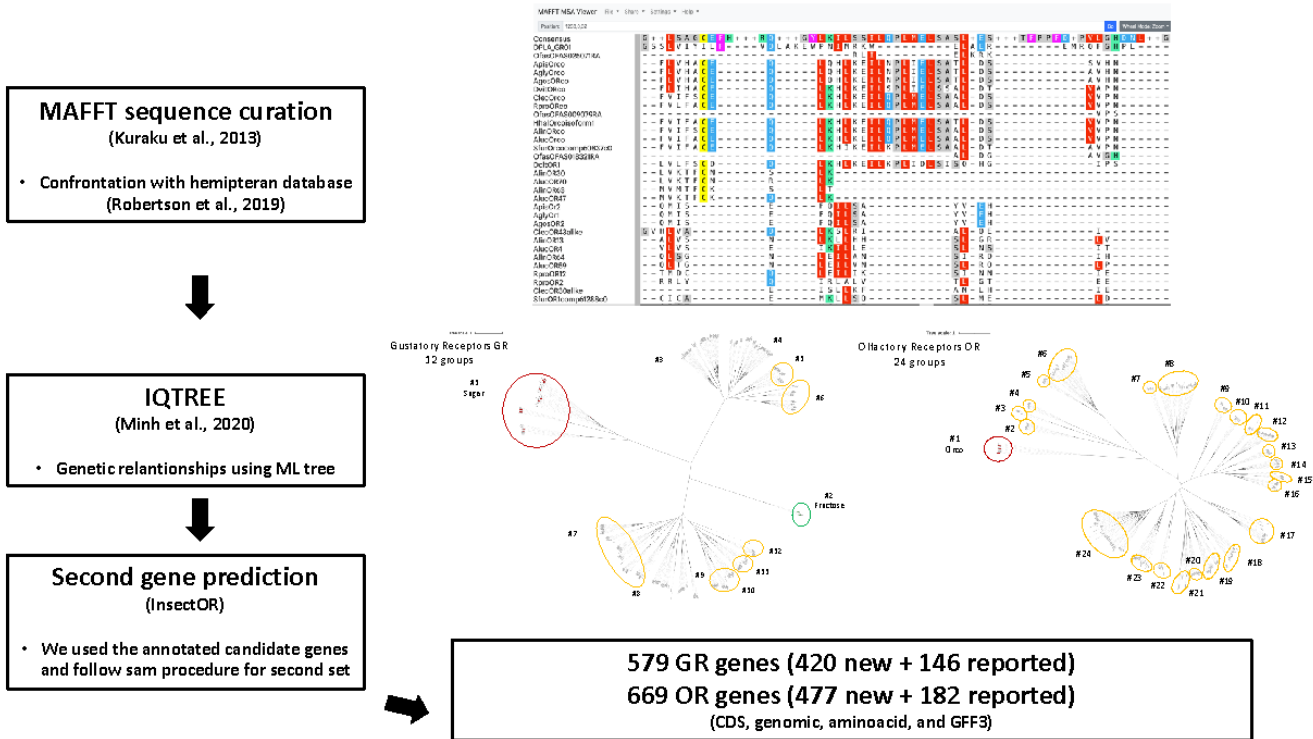
Chemosensory genes annotation



391
392
393
394
395
396
397
398
399
400
401
402
403
404

Figure S2. Chemosensory genes annotation workflow

Chemosensory genes annotation



405

406

407

408

409

410

411

412

413

414

415

416

417

418 **Figure S2 (continuation).** Chemosensory genes annotation workflow

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

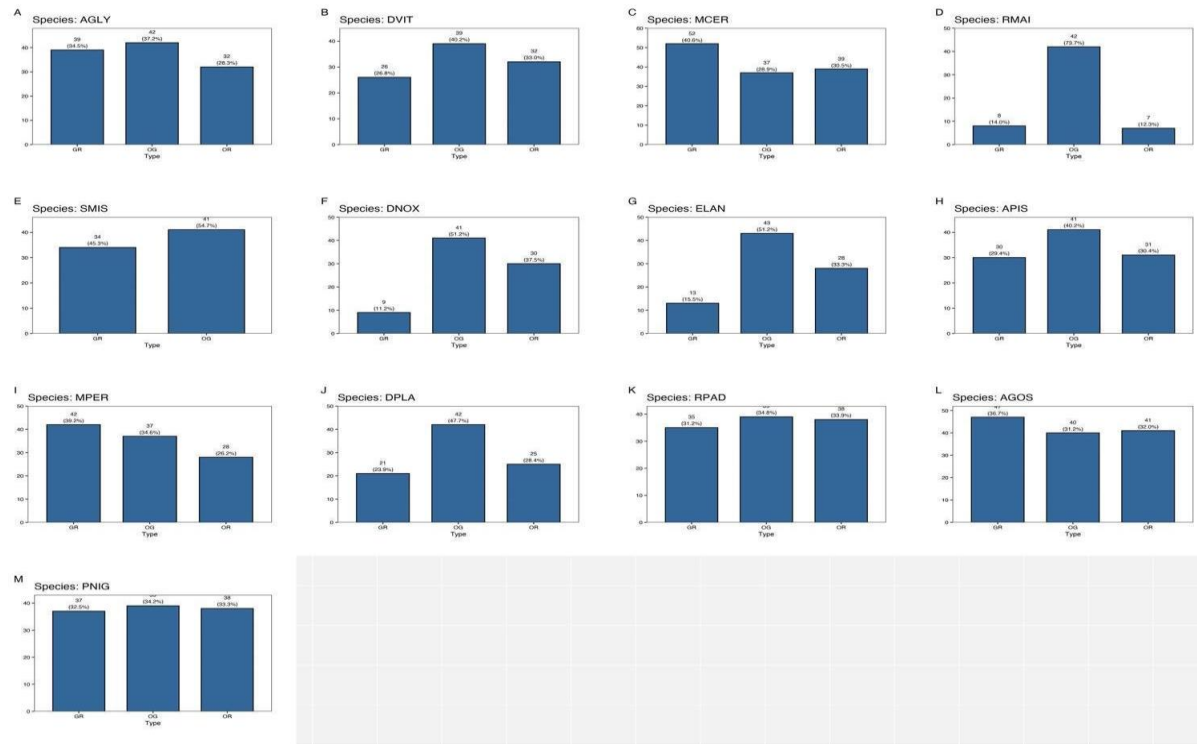


Figure S3. Number of olfactory and gustatory genes detected for the 12 aphid species and one aphid-like species, and single-copy genes used as references for the genome background. MCER: *Myzus cerasi*; MPER: *Myzus persicae*; DPLA: *Dysaphis plantaginea*; DNOX: *Diuraphis noxia*; APIS: *Acyrtosiphon pisum*; SMIS: *Sitobion miscanthi*; PNIG: *Pentalonia nigronervosa*; AGLY: *Aphis glycines*; AGOS: *Aphis gossypii*; RMAI: *Rhopalosiphum maidis*; RPAD: *Rhopalosiphum padi*; ELAN: *Eriosoma lanigerum*; DVIT: *Daktulosphaira vitifoliae*.

436 **Table S5.** Statistical test of the differences in genes numbers among species and gene family (OR and GR) annotated in the 13 aphid
437 species. Gene count was fitted to a Negative Binomial generalized linear model (glm.nb function).

438

439

440

	Df	Deviance Residuals	Pr(>Chi)
species	12	102.341	<2e-16
family (GR/OR)	1	0.137	0.7109

441

442

443 Df: degree of freedom; Pr(>Chi): P-values.

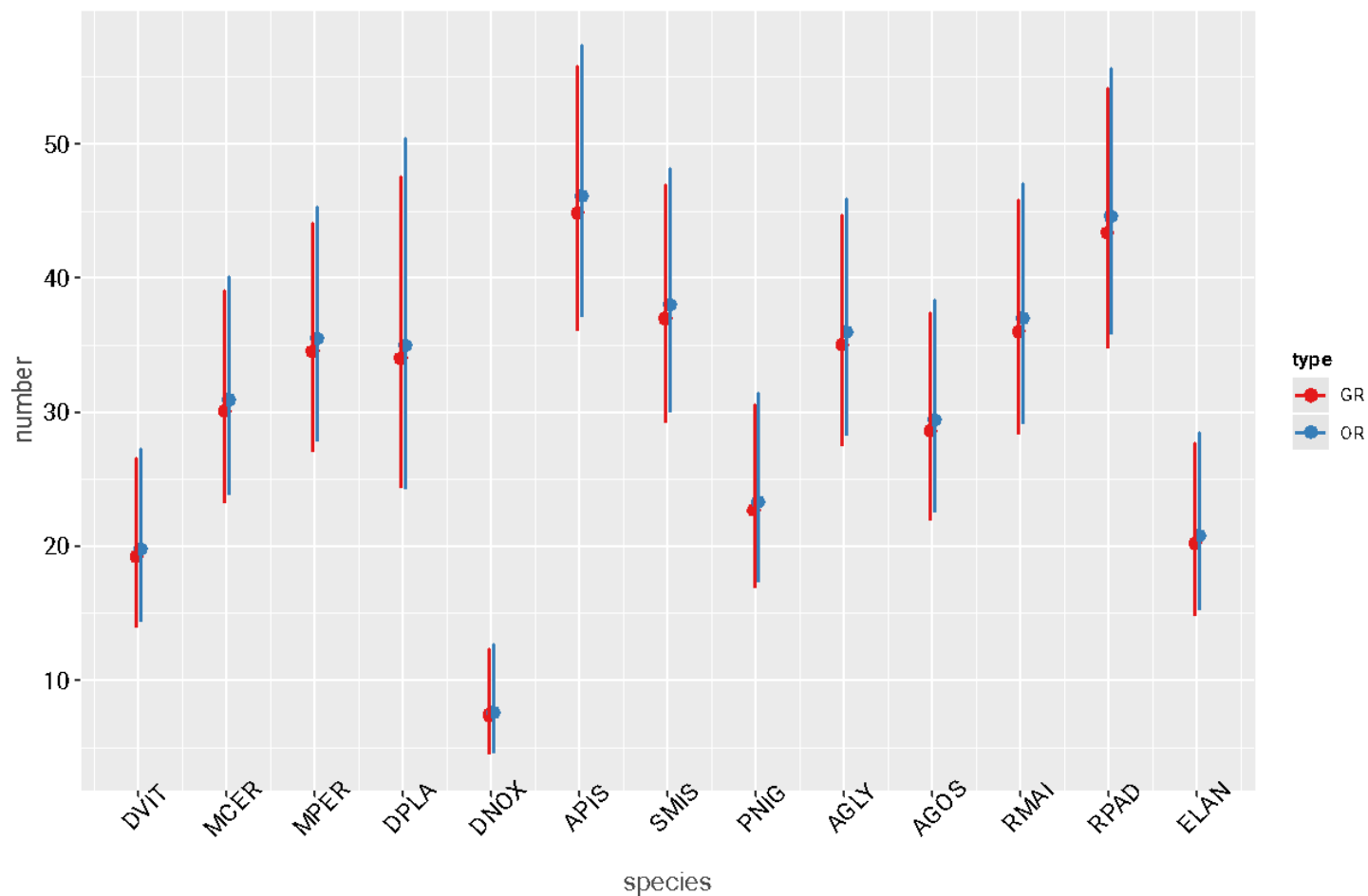
444

445

446

447

448



449

450 **Figure S4.** Number of olfactory and gustatory genes detected for the 12 aphid species and one aphid-like species predicted with
 451 Model 1 (Table S5). MCER: *Myzus cerasi*; MPER: *Myzus persicae*; DPLA: *Dysaphis plantaginea*; DNOX: *Diuraphis noxia*; APIS:
 452 *Acyrtosiphon pisum*; SMIS: *Sitobion miscanthi*; PNIG: *Pentalonia nigronervosa*; AGLY: *Aphis glycines*; AGOS: *Aphis gossypii*; RMAI:
 453 *Rhopalosiphum maidis*; RPAD: *Rhopalosiphum padi*; ELAN: *Eriosoma lanigerum*; DVIT: *Daktulosphaira vitifoliae*.

454 **Table S6.** Information about the aphid host preferences and life cycle was used in this study.

455

456

Aphid host preference information and life cycle				
Family	Subfamily	Tribe	Genera	Species
Aphididae	Aphidinae	Aphidini	<i>Aphis</i>	<i>Aphis glycines</i> (clone BT1)
Aphididae	Aphidinae	Aphidini	<i>Aphis</i>	<i>Aphis gossypii</i>
Aphididae	Aphidinae	Macrosiphini	<i>Acyrtosiphon</i>	<i>Acyrtosiphon pisum</i> (clone LSR1)
Aphididae	Aphidinae	Macrosiphini	<i>Diuraphis</i>	<i>Diuraphis noxia</i>
Aphididae	Aphidinae	Macrosiphini	<i>Dysaphis</i>	<i>Dysaphis plantaginea</i>
Phylloxeridae	Phylloxerinae	Phylloxerini	<i>Daktulosphaira</i>	<i>Daktulosphaira vitifoliae</i>
Aphididae	Eriosomatinae	Eriosomatini	<i>Eriosoma</i>	<i>Eriosoma lanigerum</i>
Aphididae	Aphidinae	Macrosiphini	<i>Myzus</i>	<i>Myzus cerasi</i>
Aphididae	Aphidinae	Macrosiphini	<i>Myzus</i>	<i>Myzus persicae</i> (clone O)
Aphididae	Aphidinae	Macrosiphini	<i>Pentalonia</i>	<i>Pentalonia nigronervosa</i>
Aphididae	Aphidinae	Aphidini	<i>Rhopalosiphum</i>	<i>Rhopalosiphum maidis</i>
Aphididae	Aphidinae	Aphidini	<i>Rhopalosiphum</i>	<i>Rhopalosiphum padi</i>
Aphididae	Aphidinae	Macrosiphini	<i>Sitobion</i>	<i>Sitobion miscanthi</i>

457

458

459

460

461 **Table S6 continuation.** Information about the aphid host preferences and life cycle used in this study.

462

463

Aphid host preference information and life cycle					
Species	Common name	Acronyms	Genome_publication	Genome Size (Mb)	Genome_assembly_origin
<i>Aphis glycines</i> (clone BT1)	Soybean aphid	AGLY	Mathers, 2020	308	USA, Illinois Urbana
<i>Aphis gossypii</i>	Cotton aphid	AGOS	Quan et al., 2019	294	Shanghai, China
<i>Acyrtosiphon pisum</i> (clone LSR1)	Pea aphid	APIS	The International Aphid Genomics Consortium, 2010	541	Itaca, New York
<i>Diuraphis noxia</i>	Russian wheat aphid	DNOX	Nicholson et al., 2015	393	Biotype2_Oklahoma_USA
<i>Dysaphis plantaginea</i>	Rosy apple aphid	DPLA	ECLECTIC-GQE-INRAE-CNRS	485	France
<i>Daktulosphaira vitifoliae</i>	Grape phylloxera	DVIT	Rispe et al., 2020	283	Bordeaux, France
<i>Eriosoma lanigerum</i>	Wolly apple aphid	ELAN	Biello et al., 2021	335	United Kingdom
<i>Myzus cerasi</i>	Black cherry aphid	MCER	Thorpe et al., 2018	406	UK
<i>Myzus persicae</i> (clone O)	Green peach aphid	MPER	Mathers et al., 2021	391	Norwich UK
<i>Pentalonia nigronervosa</i>	Bannana aphid	PNIG	Mathers et al., 2020	375	Nairobi, Kenya
<i>Rhopalosiphum maidis</i>	Maize aphid	RMAI	Chen et al., 2019	326	New York, USA
<i>Rhopalosiphum padi</i>	Bird-cherry oat aphid	RPAD	Thorpe et al., 2018	322	Henan, China
<i>Sitobion miscanthi</i>	Indian grain aphid	SMIS	Jiang et al., 2019	398	Hebei, China

464

465

466

467

468 **Table S6 continuation.** Information about the aphid host preferences and life cycle used in this study.

469

470

Acronyms	*Age	*Developmental stage	*Sex	*Host	*Tissue	*Origin	Distribution
AGLY	Adult	parthenogenic adult	Female	<i>Glycine max</i>	Whole aphid	Asia	Asia, North America, Australia, and Russia
AGOS	Adult	Adult	Female	NA	Whole aphid	Europe (tentative)	Tropical and temperate region in throughout the world
APIS	Adult	Adult	Female	<i>Medicago sativa</i>	NA	Europe or Asia	Temperate region in throughout the world
DNOX	Adult	Adult	Female	<i>Triticum aestivum</i>	Whole aphid	Central Asia	Widespread in southern Europe, Middle East, Central Asia, Africa, South and North America and Australia, but apparently not in western or northern Europe.
DPLA	Adult	Adult	Female	<i>Plantago lanceolata</i>	Whole aphid	Middle-east (under research)	Europe, Africa, much of Asia and North and South America
DVIT	Adult	Mixed leaf gall	Female	<i>Vitis vinifera</i>	Whole aphid	North America	North, Central and South America, Europe, the Mediterranean, the Middle East, Africa, China and Australia
ELAN	Adult	Adult	Female	<i>Apple plant</i>	Whole aphid	North America	Temperate region in throughout the world
MCER	Adult	Adult	NA	<i>Barbarea verna</i>	Whole aphid	Mediterranean	Temperate region in throughout the world
MPER	Adult	Single apterous female	Female	<i>Brassica rapa</i>	Whole aphid	Asia	Temperate region in throughout the world
PNIG	Adult	Adult	Female	<i>Musa sp.</i>	Whole aphid	Africa (tentative)	Tropical and subtropical region in throughout the world
RMAI	Adult	Adult	Female	<i>Zea Mays</i>	Whole aphid	Asia	Tropical and subtropical region in throughout the world
RPAD	Adult	Adult	Female	<i>Triticum aestivum</i>	Whole aphid	North America	Temperate region in throughout the world
SMIS	Adult	Newborn nymph	Female	<i>Triticum aestivum</i>	Whole aphid	Asian	Indian subcontinent, east and south-east Asia and Australasia

471

472 **Table S6 continuation.** Information about the aphid host preferences and life cycle used in this study.

Acronyms	Life cycle	Reproduction	Host specialization	Host_primary	Host_secondary
AGLY	Heteroecious	Holocyclic	biotypes reported	<i>Rhamnus</i> sp.	Mainly <i>Glycine max</i> (soybean) but also <i>Glycine</i> sp. and other Fabaceae members
AGOS	Monoecious	Holocyclic/Anholocyclic in tropics regions	biotypes reported	Catalpa bignonioides (Indian bean tree), Hibiscus syriacus (Korean rose), Celastrus orbiculatus (oriental bittersweet), <i>Rhamnus</i> species (buckthorns) and <i>Punica granatum</i> (pomegranate)	Cucurbitaceae, Malvaceae, and Rutaceae species
APIS	Monoecious	Holocyclic/Anholocyclic in mild climates	biotypes reported	Family Fabaceae, but especially on <i>Medicago</i> , <i>Melilotus</i> , <i>Trifolium</i> , <i>Dorycnium</i> and <i>Lotus</i>	
DNOX	Monoecious	Holocyclic/mainly Anholocyclic	biotypes reported	Grasses and cereals (such as <i>Agropyron</i> , <i>Bromus</i> , <i>Elymus</i> , <i>Hordeum</i> , <i>Triticum</i>)	
DPLA	Dioecious	Holocyclic	not reported	<i>Malus domestica</i> (cultivated apple)	<i>Plantago lanceolata</i> (ribwort plantain)
DVIT	Monoecious	Anholocyclic	not reported	<i>Vitis</i> sp. (grape)	
ELAN	Dioecious in America/Monoecious most part of the world	Anholocyclic/Holocyclic (in America)	not reported	<i>Ulmus americana</i> (Elm) in America / <i>Malus domestica</i> (other parts of the world)	<i>Malus domestica</i> (cultivated apple)
MCER	Dioecious	Holocyclic	not reported	<i>Prunus cerasus/Prunus avium</i> (cherry trees)	bedstraws (<i>Galium</i>), eyebrights (<i>Euphrasia</i>) and speedwell (<i>Veronica</i> sp.) Brassicaceae, Solanaceae, Poaceae, Leguminosae, Cypraceae, Convolvulaceae, Chenopodiaceae, Compositae, Cucurbitaceae, and Umbelliferae
MPER	Heteroecious	Holocyclic/Anholocyclic in <i>M. persicae nicotiana</i>	biotypes reported	<i>Prunus persica</i> (peach)	
PNIG	Monoecious	Anholocyclic	not reported	Musaceae	
RMAI	Monoecious/Dioecious in Asia	Anholocyclic/Holocyclic (in Asia)	biotypes reported	Poaceae/ <i>Prunus</i> sp. (Asia)	Poaceae
RPAD	Dioecious	Holocyclic/Anholocyclic in mild climates	not reported	<i>Prunus padus</i> (bird cherry)	Poaceae
SMIS	Monoecious	Anholocyclic	biotypes reported		

473
474 Holocyclic: reproduce both asexually and sexually; Anholocyclic: asexual reproduction only; Monoecious: same host throughout their life stages; Dioecious/Heteroecious: host alternates between sexual and asexual stages (Hardy et al., 2015; Emden and Harrington, 2017).

475 **Table S7.** Likelihood-ratio test of models with different evolutionary rates of the 12 aphid and one aphid-like species depending on
 476 the taxonomic ranks (genus, subfamilies, families, species) or host preferences (see details in Table S5) or with a single evolutionary
 477 rate along the phylogeny inferred with CAFE (Han et al., 2013).

478

Model	P vs. S	F vs.S	Sf vs. S	T vs. S	G vs. S	F vs. P	Sf vs.s P	T vs. P	P vs. G	F vs. Sf	F vs. T	F vs. G	Sf vs. T	Sf vs. G	T vs. G	P vs. S	S vs. Gr	S vs. H	P vs. Sp	P vs. H	H vs. Gr
Ho (InL)	254806	247150	243475	244584	266625	247150	243475	244584	254806	247150	247150	247150	243475	243475	244584	267684	255490	267678	255490	267678	255490
Ha (InL)	255636	255636	255636	255636	255636	254806	254806	254806	266625	243475	244584	266625	244584	266625	266625	284810	255490	267684	284810	284810	267678
2ΔL	1660	16972	24322	22104	-21978	-15312	-22662	-20444	23638	7350	5132	-38950	-2218	-46300	-44082	34252	24388	12	58640	34264	24376
K	5	1	2	3	9	7	6	5	1	1	2	8	1	7	6	1	1	1	1	1	1
alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500
p-value	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	1	1	1	1	<2.2e-16	<2.2e-16	<2.2e-16	1	1	1	1	2.2e-16	2.2e-16	0.000532	2.2e-16	2.2e-16	2.2e-16
χ²	11.07	3.8415	5.9915	7.8147	16.919	14.067	12.592	11.07	3.8415	3.8415	5.9915	15.507	3.8415	14.067	12.592	3.8415	3.8415	3.8415	3.8415	3.8415	3.8415
Best hypothesis	P	F	Sf	T	S	F	Sf	T	G	Sf	T	F	Sf	Sf	T	P	Gr	H	P	P	H

479

480 **Ho (InL) : Nule hypothesis ; Ha (InL) : Alternative hypothesis; K :Degrees of freedom; alpha : 0.05 ; p-value : <0.05 ; χ : chi-square distribution; Hypotheses acronyms P=Phylogeny, S=single, F=Family, Sf=Subfamily, T=Tribe,**
 481 **G=Genus, H=Host, Gr=Generalist, Sp : Specialist**

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

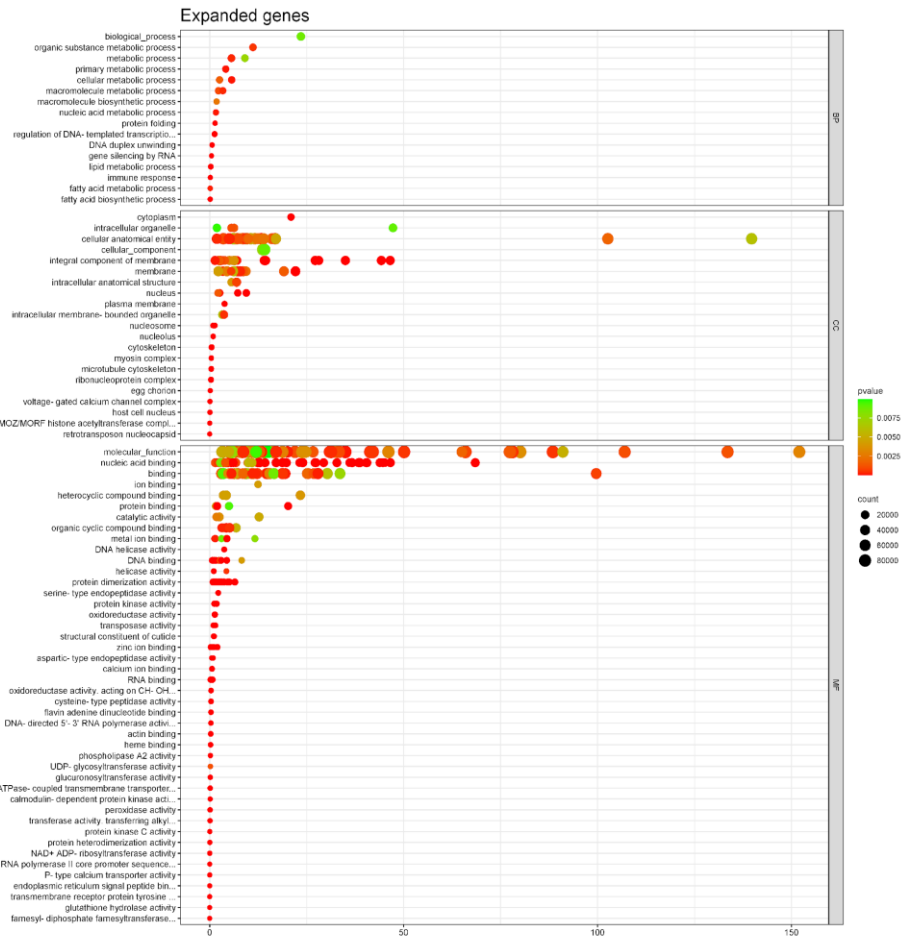


Figure S5. Enrichment analysis of genes underwent expansion during the aphid and aphid-like species.

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

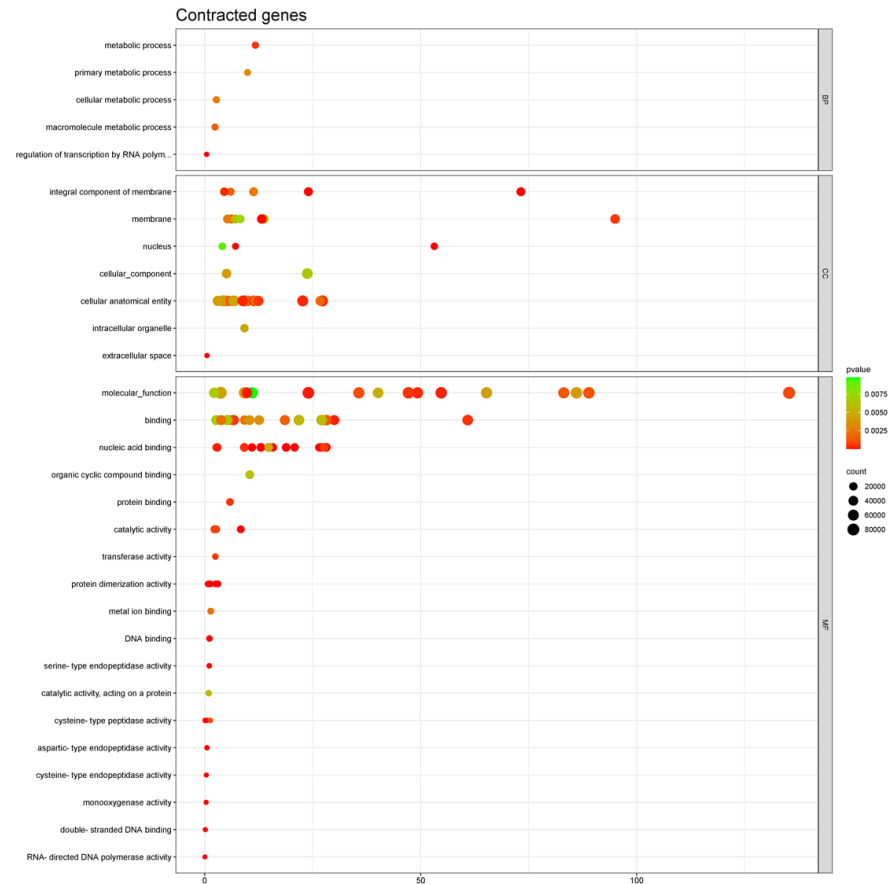


Figure S6. Enrichment analysis of genes underwent contraction during the aphid and aphid-like species.

515 **Table S8.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000003	GO:0003676	nucleic acid binding	25443	220	68.47	< 1e-30	MF
	GO:0110165	cellular anatomical entity	61766	10	9.59	0.0030	CC
OG0000010	GO:0046872	metal ion binding	6937	7	4.27	0.0062	MF
OG0000011	GO:0046983	protein dimerization activity	6236	151	6.5	< 1e-30	MF
	GO:0003674	molecular_function	84894	181	88.5	0.00052	MF
	GO:0016020	membrane	36345	9	7.12	0.00015	CC
	GO:0043231	intracellular membrane-bounded organelle	16414	6	3.22	0.00632	CC
OG0000013			71272	13	14.04	0.00015	MF
OG0000015	GO:0003674	molecular_function	59329	23	14.78	0.0026	MF
	GO:0005488	binding	39720	17	9.9	0.0082	MF
	GO:0110165	cellular anatomical entity	38131	12	13.87	0.0005	CC
	GO:0016021	integral component of membrane	17588	6	6.4	0.0011	CC
OG0000016	O:0003676	nucleic acid binding	24203	108	38.4	<1e-30	MF
	GO:0004190	aspartic-type endopeptidase activity	446	25	0.71	<1e-30	MF
OG0000018	GO:0032508	DNA duplex unwinding	320	34	0.64	< 1e-30	BP
	GO:0003678	DNA helicase activity	998	87	3.76	< 1e-30	MF
	GO:0004386	helicase activity	1145	89	4.32	0.00055	MF
OG0000019	GO:0042302	structural constituent of cuticle	808	171	1.14	< 1e-30	MF
	GO:0005730	nucleolus	522	22	0.9	< 1e-30	CC
OG0000023	GO:0003676	nucleic acid binding	23945	9	6.74	0.00087	MF
	GO:0003674	molecular_function	78123	19	22	0.00104	MF
OG0000025	GO:0046983	protein dimerization activity	6075	110	4.87	< 1e-30	MF
	GO:0110165	cellular anatomical entity	54563	9	10.63	0.00094	CC

517 **Table S8 continuation.** Enrichment analysis of expanded genes.

518

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000026	GO:0003677	DNA binding	4665	65	4.4	< 1e-30	MF
	GO:0003676	nucleic acid binding	25443	143	24	< 1e-30	MF
	GO:0003674	molecular_function	84894	160	80.08	0.0035	MF
	GO:0110165	cellular anatomical entity	61766	17	10.59	0.0053	CC
	GO:0016020	membrane	36345	7	6.23	0.0078	CC
OG0000028	GO:0004803	transposase activity	762	64	1.5	< 1e-30	MF
	GO:0003676	nucleic acid binding	22686	101	44.69	< 1e-30	MF
	GO:0110165	cellular anatomical entity	50745	13	16.05	0.00179	CC
	GO:0016020	membrane	29607	8	9.36	0.00230	CC
OG0000029	GO:0003676	nucleic acid binding	25443	160	36.71	< 1e-30	MF
OG0000030	GO:0046983	protein dimerization activity	6236	74	3.14	< 1e-30	MF
	GO:0110165	cellular anatomical entity	61766	17	9.59	0.0018	CC
OG0000031	GO:0003674	molecular_function	58064	17	20.07	0.0002	MF
OG0000032	GO:0003674	molecular_function	84894	7	5.05	0.00145	MF
	GO:0110165	cellular anatomical entity	61766	7	7.06	0.00403	CC
OG0000034	O:0046872	metal ion binding	7322	33	1.45	<1e-30	MF
	GO:0005488	binding	55053	38	10.91	0.0014	MF
	GO:0016021	integral component of membrane	28001	5	2.63	0.00081	CC
OG0000035	GO:0071704	organic substance metabolic process	10041	16	11.16	0.00057	BP
	GO:0005515	protein binding	14711	80	20.23	< 1e-30	MF
OG0000036	GO:0005488	binding	55053	19	7.2	0.00032	MF
	GO:0110165	cellular anatomical entity	61766	27	9.33	0.00022	CC

519 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000038	GO:0003676	nucleic acid binding	20152	108	45.09	< 1e-30	MF
	GO:0005634	nucleus	7961	5	2.51	0.00083	CC
	GO:0005622	intracellular anatomical structure	17867	8	5.63	0.00410	CC
	GO:0110165	cellular anatomical entity	42246	11	13.32	0.00427	CC
OG0000039	GO:0003676	nucleic acid binding	25443	81	38.72	< 1e-30	MF
	GO:0004803	transposase activity	767	26	1.17	< 1e-30	MF
OG0000040	GO:0003676	nucleic acid binding	25443	8	5.55	0.00065	MF
	GO:0005634	nucleus	11058	6	2.57	0.00018	CC
	GO:0016021	integral component of membrane	28001	6	6.52	0.00296	CC
OG0000042	GO:0003676	nucleic acid binding	20851	155	36.11	<1e-30	MF
OG0000045	GO:0003676	nucleic acid binding	22097	96	43.89	<1e-30	MF
	GO:0004803	transposase activity	702	34	1.39	<1e-30	MF
	GO:0003677	DNA binding	4158	8	8.26	0.0042	MF
	GO:0110165	cellular anatomical entity	50037	5	6.6	0.0011	CC
OG0000046	GO:0016021	integral component of membrane	22139	7	6.08	0.00081	CC
	GO:0110165	cellular anatomical entity	48111	15	13.21	0.00173	CC
	GO:0016020	membrane	28700	10	7.88	0.00291	CC
OG0000049	GO:0046983	protein dimerization activity	6075	110	4.69	< 1e-30	MF
	GO:0110165	cellular anatomical entity	54563	6	5.82	0.00026	CC
OG0000050	GO:0003676	nucleic acid binding	25443	63	19.77	< 1e-30	MF
	GO:0003674	molecular_function	84894	79	65.95	0.00088	MF
	GO:1901363	heterocyclic compound binding	30095	66	23.38	0.00395	MF

520

521 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000053	GO:0110165	cellular anatomical entity	51496	5	6.01	0.0039	CC
OG0000054	GO:0003674	molecular_function	78861	18	12.31	0.00102	MF
OG0000057	GO:0005488	binding	47581	11	5.25	0.00012	MF
OG0000058	GO:0003676	nucleic acid binding	25443	6	4.24	0.0015	MF
	GO:0003674	molecular_function	84894	14	14.13	0.0093	MF
OG0000059	GO:0008150	biological_process	19881	14	23.52	0.0087	BP
	GO:0015020	glucuronosyltransferase activity	247	30	0.16	< 1e-30	MF
	GO:0008194	UDP-glycosyltransferase activity	289	31	0.18	0.0012	MF
	GO:0016021	integral component of membrane	28001	109	46.54	< 1e-30	CC
	GO:0110165	cellular anatomical entity	61766	124	102.66	0.0018	CC
OG0000060	GO:0008270	zinc ion binding	2161	94	1.99	< 1e-30	MF
	GO:0110165	cellular anatomical entity	57948	5	6.5	0.0038	CC
OG0000061	GO:0008152	metabolic process	11177	12	9.08	0.0074	BP
	GO:0004252	serine-type endopeptidase activity	982	91	2.21	< 1e-30	MF
	GO:0016020	membrane	36345	8	8.61	0.0015	CC
OG0000063	GO:0046983	protein dimerization activity	6236	81	5.04	< 1e-30	MF
	GO:0110165	cellular anatomical entity	61766	8	9.33	0.0065	CC
OG0000064	GO:0003676	nucleic acid binding	25443	5	2.62	0.00956	MF
OG0000065	GO:0046872	metal ion binding	7322	33	1.42	< 1e-30	MF
	GO:0003674	molecular_function	84894	39	16.49	0.0032	MF
	GO:0110165	cellular anatomical entity	61766	7	8.58	0.00025	CC
OG0000066	GO:0003676	nucleic acid binding	24538	116	40.54	< 1e-30	MF
	GO:0003674	molecular_function	80821	127	133.54	0.0012	MF
OG0000068	O:0003674	molecular_function	71704	12	8.9	0.00032	MF
	GO:0003824	catalytic activity	20039	6	2.49	0.00936	MF

522 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000069	GO:0003779	actin binding	614	27	0.27	<1e-30	MF
	GO:0016459	myosin complex	304	88	0.43	<1e-30	CC
OG0000070	GO:0004190	aspartic-type endopeptidase activity	450	30	0.88	< 1e-30	MF
	GO:0003676	nucleic acid binding	23945	81	46.59	< 1e-30	MF
	GO:0005488	binding	51222	96	99.67	0.00078	MF
	GO:0003674	molecular_function	78123	129	152.02	0.00304	MF
OG0000071	GO:0110165	cellular anatomical entity	61766	16	13.37	0.00014	CC
	GO:0016020	membrane	36345	9	7.87	0.00027	CC
	GO:0005575	cellular_component	63112	17	13.66	0.00735	CC
OG0000072	GO:0003676	nucleic acid binding	23945	71	32.8	<1e-30	MF
	GO:0004803	transposase activity	766	28	1.05	<1e-30	MF
	GO:0003674	molecular_function	78123	104	107.01	0.0011	MF
OG0000073	GO:0046983	protein dimerization activity	6236	97	3.68	< 1e-30	MF
	GO:0003674	molecular_function	84894	122	50.14	0.00052	MF
OG0000074	GO:0006355	regulation of DNA-templated transcriptio...	1915	42	1.27	< 1e-30	BP
	GO:0044237	cellular metabolic process	8579	50	5.67	0.00012	BP
	GO:0009059	macromolecule biosynthetic process	2695	44	1.78	0.00248	BP
OG0000075	GO:0003674	molecular_function	72090	8	6.58	0.0038	MF
	GO:0110165	cellular anatomical entity	50037	11	7.37	0.00051	CC
OG0000077	GO:0043170	macromolecule metabolic process	7535	8	2.28	0.00099	BP
	GO:0044237	cellular metabolic process	8474	8	2.56	0.00163	BP
OG0000078	GO:0005886	plasma membrane	1859	113	3.84	<1e-30	CC
OG0000079	GO:0003676	nucleic acid binding	23945	73	9.91	<1e-30	MF
OG0000080	GO:0003676	nucleic acid binding	23945	50	27.18	< 1e-30	MF

523 **Table S8 continuation.** Enrichment analysis of expanded genes.

524

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000082	GO:0046983	protein dimerization activity	6236	77	2.57	< 1e-30	MF
	GO:0003674	molecular_function	84894	85	34.99	0.00011	MF
OG0000085	GO:0005488	binding	51222	20	16.61	0.0061	MF
OG0000093	GO:0005488	binding	55053	7	3.05	0.0014	MF
OG0000094	GO:0003676	nucleic acid binding	22495	65	40.34	< 1e-30	MF
	GO:0046872	metal ion binding	6505	9	11.67	0.0078	MF
	GO:0110165	cellular anatomical entity	48111	6	7.87	0.00410	CC
OG0000097	GO:0003674	molecular_function	72090	53	30.94	0.00042	MF
	GO:0005634	nucleus	8629	5	2.15	0.0014	CC
	GO:0016021	integral component of membrane	23278	6	5.79	0.0022	CC
OG0000099	GO:0004623	phospholipase A2 activity	145	40	0.19	<1e-30	MF
	GO:0003674	molecular_function	50861	47	65.02	0.0018	MF
OG0000100	GO:0003674	molecular_function	78475	25	15	0.00993	MF
OG0000101	GO:0046983	protein dimerization activity	5843	50	2.79	< 1e-30	MF
OG0000103	GO:0031047	gene silencing by RNA	78	20	0.49	<1e-30	BP
	GO:0000979	RNA polymerase II core promoter sequence...	10	9	0.02	< 1e-30	MF
	GO:0000786	nucleosome	602	86	1.36	< 1e-30	CC
	GO:0110165	cellular anatomical entity	61766	97	139.74	0.0061	CC
	GO:0043229	intracellular organelle	20891	94	47.27	0.0090	CC
OG0000104	GO:0003676	nucleic acid binding	22495	65	29.38	< 1e-30	MF
	GO:0004803	transposase activity	753	24	0.98	< 1e-30	MF
OG0000107	GO:0003674	molecular_function	71704	12	9.89	0.0077	MF

525 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000108	GO:0003676	nucleic acid binding	25443	13	12.4	0.00046	MF
	GO:0003674	molecular_function	84894	46	41.39	0.00052	MF
OG0000110	GO:0005488	binding	55053	9	3.49	0.00931	MF
	GO:0110165	cellular anatomical entity	61766	6	3.53	0.0010	CC
OG0000113	GO:0003674	molecular_function	78123	22	11	0.0032	MF
OG0000118	O:0110165	cellular anatomical entity	54563	14	8.35	0.00178	CC
OG0000122	GO:0110165	cellular anatomical entity	61766	16	16.9	0.00076	CC
	GO:0005622	intracellular anatomical structure	25380	12	6.94	0.00080	CC
OG0000123	GO:0046983	protein dimerization activity	5937	47	2.2	< 1e-30	MF
OG0000124	GO:0005488	binding	51412	13	6.99	0.0029	MF
	GO:0016020	membrane	32671	5	6.12	0.0055	CC
OG0000126	GO:0046983	protein dimerization activity	6075	56	2.36	< 1e-30	MF
OG0000127	GO:0003674	molecular_function	78475	46	32.67	0.0032	MF
	GO:0003676	nucleic acid binding	23993	11	9.99	0.0056	MF
OG0000130	GO:0003676	nucleic acid binding	18894	55	23.54	<1e-30	MF
	GO:0004803	transposase activity	743	21	0.93	<1e-30	MF
OG0000132	GO:0003899	DNA-directed 5'-3' RNA polymerase activi...	339	19	0.31	< 1e-30	MF
	GO:0003674	molecular_function	84894	26	78.06	0.00048	MF
	GO:0000786	nucleosome	602	96	1.04	<1e-30	CC
OG0000137	GO:0005488	binding	46385	20	8.91	0.0026	MF
OG0000138	GO:0005488	binding	51412	16	6.33	0.00061	MF
OG0000140	GO:0003676	nucleic acid binding	25443	81	17.14	< 1e-30	MF
OG0000141	GO:0046983	protein dimerization activity	6075	45	1.74	<1e-30	MF
	GO:0005488	binding	51222	49	14.64	0.0015	MF

526 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000142	GO:0046983	protein dimerization activity	6075	53	2.59	< 1e-30	MF
	GO:0003674	molecular_function	78123	60	33.34	0.00099	MF
	GO:0110165	cellular anatomical entity	54563	6	9.37	0.00026	CC
OG0000147	GO:0003676	nucleic acid binding	23595	75	27.38	< 1e-30	MF
OG0000150	GO:0003674	molecular_function	84894	14	11.78	0.0042	MF
	GO:0003676	nucleic acid binding	25443	5	3.53	0.0096	MF
	GO:0110165	cellular anatomical entity	61766	5	6.31	0.00404	CC
OG0000151	GO:0046982	protein heterodimerization activity	144	11	0.04	< 1e-30	MF
	GO:0000786	nucleosome	602	112	1.26	< 1e-30	CC
OG0000152	GO:0003677	DNA binding	4408	35	2.37	< 1e-30	MF
	GO:0003676	nucleic acid binding	23945	44	12.87	0.0078	MF
	GO:0043229	intracellular organelle	18086	5	1.85	0.0097	CC
OG0000153	GO:0003676	nucleic acid binding	22686	6	3.58	0.0019	MF
OG0000155	GO:0016021	integral component of membrane	23278	7	5.08	0.00029	CC
OG0000159	GO:0003676	nucleic acid binding	25443	72	18.76	< 1e-30	MF
OG0000161	GO:0004386	helicase activity	1163	25	1.06	< 1e-30	MF
	GO:0003674	molecular_function	84894	31	77.39	0.0014	MF
	GO:0016021	integral component of membrane	28001	86	34.99	< 1e-30	CC
OG0000167	GO:0005488	binding	46181	21	15.82	0.0015	MF
OG0000168	GO:0046983	protein dimerization activity	6236	65	2.3	< 1e-30	MF
	GO:0003674	molecular_function	84894	72	31.29	0.00034	MF
	GO:0110165	cellular anatomical entity	61766	13	11.6	0.0040	CC
OG0000169	GO:0046983	protein dimerization activity	6075	62	3.01	< 1e-30	MF
	GO:0005488	binding	51222	71	25.36	0.0014	MF

527 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000171	GO:0003674	molecular_function	66918	10	13.36	0.00015	MF
OG0000178	GO:0016021	integral component of membrane	22491	5	3.22	0.00071	CC
OG0000185	GO:0110165	cellular anatomical entity	41187	6	1.83	0.00032	CC
OG0000186	GO:0008270	zinc ion binding	1874	71	1.02	<1e-30	MF
	GO:0005488	binding	45991	78	25.12	0.0018	MF
	GO:0046872	metal ion binding	5569	73	3.04	0.0088	MF
OG0000189	GO:0000943	retrotransposon nucleocapsid	13	9	0	<1e-30	CC
OG0000191	O:0046983	protein dimerization activity	6075	32	2.31	< 1e-30	MF
	GO:0005488	binding	51222	44	19.45	0.00017	MF
OG0000192	GO:0046983	protein dimerization activity	6075	42	2.28	< 1e-30	MF
	GO:0005515	protein binding	13118	45	4.93	0.004	MF
	GO:0016021	integral component of membrane	24921	5	5.09	0.0031	CC
OG0000194	GO:0046983	protein dimerization activity	6079	42	2.22	< 1e-30	MF
OG0000196	GO:0046872	metal ion binding	5184	5	1.28	0.00023	MF
	GO:0003676	nucleic acid binding	20152	7	4.97	0.00508	MF
	GO:0016021	integral component of membrane	18242	5	3.91	0.0013	CC
OG0000197	GO:0003824	catalytic activity	22400	5	2.49	0.0036	MF
OG0000198	GO:0016021	integral component of membrane	18892	8	7.07	0.00013	CC
	GO:0043229	intracellular organelle	16817	8	6.29	0.00136	CC
	GO:0110165	cellular anatomical entity	45630	17	17.08	0.00476	CC
OG0000204	GO:0043231	intracellular membrane-bounded organelle	16414	8	3.69	0.00044	CC
	GO:0016021	integral component of membrane	28001	7	6.29	0.00369	CC
	GO:0005575	cellular_component	63112	19	14.18	0.00899	CC

529 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000220	GO:0044238	primary metabolic process	7506	7	4.15	0.00011	BP
	GO:0043170	macromolecule metabolic process	6162	6	3.41	0.00034	BP
	GO:0004190	aspartic-type endopeptidase activity	396	30	0.57	< 1e-30	MF
	GO:0003674	molecular_function	63067	62	91.07	0.0052	MF
OG0000222	GO:0003676	nucleic acid binding	22495	5	1.45	0.00104	MF
OG0000224	GO:0046983	protein dimerization activity	6236	42	1.68	<1e-30	MF
	GO:0005488	binding	55053	52	14.84	0.0002	MF
	GO:0003677	DNA binding	4665	5	1.26	0.0027	MF
OG0000230	GO:0005488	binding	55053	25	9.38	0.0002	MF
	GO:0003676	nucleic acid binding	25443	20	4.34	0.0069	MF
OG0000231	GO:0006457	protein folding	319	18	1.39	< 1e-30	BP
	GO:0005737	cytoplasm	9440	72	20.97	< 1e-30	CC
OG0000235	GO:0046872	metal ion binding	7322	22	4.47	<1e-30	MF
	GO:0005488	binding	55053	25	33.6	0.0074	MF
	GO:0000786	nucleosome	602	72	0.85	<1e-30	CC
OG0000237	GO:0042626	ATPase-coupled transmembrane transporter...	288	20	0.14	< 1e-30	MF
	GO:0016021	integral component of membrane	28001	64	44.25	<1e-30	CC
OG0000238	GO:0046983	protein dimerization activity	5995	62	2.89	< 1e-30	MF
OG0000240	GO:0008152	metabolic process	11177	7	5.61	0.00032	BP
	GO:0003676	nucleic acid binding	25443	5	2.52	0.0029	MF
	GO:0003674	molecular_function	84894	11	8.41	0.0047	MF
OG0000242	GO:0003674	molecular_function	72951	10	7.78	0.0044	MF
OG0000243	O:0046983	protein dimerization activity	6236	31	0.91	<1e-30	MF

530

531 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000244	GO:0003674	molecular_function	71254	15	11.25	0.00030	MF
	GO:0016021	integral component of membrane	21674	6	2.57	0.00063	CC
	GO:0016020	membrane	29208	9	3.47	0.00116	CC
OG0000246	GO:0036374	glutathione hydrolase activity	9	9	0	< 1e-30	MF
	GO:0016021	integral component of membrane	25880	41	28.14	< 1e-30	CC
OG0000250	GO:0003676	nucleic acid binding	25443	45	12.61	< 1e-30	MF
	GO:0003674	molecular_function	84894	58	42.06	0.00052	MF
OG0000256	GO:0003677	DNA binding	4158	52	1.63	< 1e-30	MF
	GO:0043229	intracellular organelle	16275	14	5.62	0.0011	CC
OG0000259	GO:0005488	binding	39203	26	16.36	0.0091	MF
	GO:0016020	membrane	21563	6	3.06	0.00042	CC
OG0000260	GO:0003677	DNA binding	4149	30	1.57	< 1e-30	MF
	GO:0110165	cellular anatomical entity	52859	16	12.34	0.0022	CC
OG0000264	GO:1901363	heterocyclic compound binding	24847	7	3.5	0.00427	MF
	GO:0097159	organic cyclic compound binding	24955	7	3.51	0.00435	MF
OG0000278	GO:0046983	protein dimerization activity	6236	47	1.75	< 1e-30	MF
	GO:0005488	binding	55053	53	15.49	0.0029	MF
	GO:0003674	molecular_function	84894	56	23.89	0.0035	MF
	GO:0016020	membrane	36345	5	4.45	0.0027	CC
OG0000285	GO:0003676	nucleic acid binding	25443	43	4.64	< 1e-30	MF
OG0000287	GO:0004601	peroxidase activity	209	13	0.11	< 1e-30	MF
	GO:0003674	molecular_function	84894	23	46.1	0.0031	MF
	GO:0042600	egg chorion	122	40	0.16	< 1e-30	CC
OG0000292	GO:0016021	integral component of membrane	28001	55	14.07	< 1e-30	CC
OG0000297	GO:0046983	protein dimerization activity	6079	45	2.25	< 1e-30	MF

532 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000298	GO:0008270	zinc ion binding	2499	20	0.43	< 1e-30	MF
	GO:0005488	binding	55053	25	9.38	0.002	MF
	GO:0016021	integral component of membrane	28001	5	2.74	0.00081	CC
	GO:0110165	cellular anatomical entity	61766	8	6.05	0.00377	CC
OG0000305	GO:0003676	nucleic acid binding	19458	5	2.16	0.00049	MF
OG0000307	GO:0003677	DNA binding	4486	28	0.67	< 1e-30	MF
OG0000308	GO:0016020	membrane	36345	6	2.23	0.0043	CC
OG0000310	GO:0003674	molecular_function	78475	21	10.67	0.0031	MF
OG0000311	GO:0003676	nucleic acid binding	17445	43	23.17	< 1e-30	MF
OG0000320	GO:0003950	NAD+ ADP-ribosyltransferase activity	84	32	0.03	< 1e-30	MF
OG0000324	GO:0004697	protein kinase C activity	38	26	0.04	< 1e-30	MF
	GO:0043167	ion binding	10596	7	12.47	0.0042	MF
OG0000325	GO:0015630	microtubule cytoskeleton	941	51	0.41	< 1e-30	CC
OG0000326	GO:0003674	molecular_function	71704	9	6.59	0.00127	MF
OG0000333	GO:0003674	molecular_function	78475	5	3.33	0.0041	MF
OG0000334	GO:0003677	DNA binding	4408	5	2.8	0.0029	MF
	GO:0110165	cellular anatomical entity	54563	6	5.57	0.008	CC
OG0000337	GO:0046983	protein dimerization activity	6236	42	1.43	< 1e-30	MF
OG0000338	GO:0003674	molecular_function	58064	6	5.19	0.0017	MF
OG0000340	GO:0005488	binding	47581	31	26.91	0.0012	MF
OG0000342	GO:0004683	calmodulin-dependent protein kinase acti...	86	56	0.13	< 1e-30	MF
OG0000347	GO:0046983	protein dimerization activity	6236	41	2.3	< 1e-30	MF
OG0000351	GO:0003674	molecular_function	76424	16	12.78	0.0014	MF
OG0000354	GO:0003676	nucleic acid binding	18894	10	5.29	0.00031	MF
	GO:0110165	cellular anatomical entity	38427	6	6.94	0.0037	CC

533 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000356	GO:0005891	voltage-gated calcium channel complex	101	50	0.09	<1e-30	CC
OG0000361	GO:0003674	molecular_function	84894	16	8.41	0.00315	MF
OG0000371	GO:0005488	binding	55053	36	11.13	0.0022	MF
OG0000377	GO:0003674	molecular_function	58064	25	26.65	0.0013	MF
OG0000378	GO:0005634	nucleus	11058	33	9.44	<1e-30	CC
OG0000383	GO:0005488	binding	55053	24	30.33	0.0061	MF
	GO:0070776	MOZ/MORF histone acetyltransferase compl...	35	20	0.03	< 1e-30	CC
OG0000384	GO:0003674	molecular_function	84894	12	5.72	0.0076	MF
	GO:0042025	host cell nucleus	156	27	0.08	< 1e-30	CC
OG0000385	GO:0004672	protein kinase activity	2357	50	1.79	<1e-30	MF
OG0000386	GO:0046983	protein dimerization activity	6156	31	1.21	<1e-30	MF
OG0000401	GO:0005509	calcium ion binding	1112	35	0.61	<1e-30	MF
OG0000403	GO:0003676	nucleic acid binding	25443	41	12.91	<1e-30	MF
	GO:0005488	binding	55053	46	27.93	0.0002	MF
OG0000408	GO:0042302	structural constituent of cuticle	808	43	1.01	< 1e-30	MF
OG0000409	GO:0005488	binding	44389	8	6.42	0.0047	MF
OG0000411	GO:0046983	protein dimerization activity	4462	5	1.08	0.00054	MF
	GO:0110165	cellular anatomical entity	21659	7	6	0.00311	CC
OG0000414	GO:0003674	molecular_function	58064	5	3.11	0.005	MF
OG0000415	GO:0016491	oxidoreductase activity	3049	34	1.33	< 1e-30	MF
OG0000417	GO:0003676	nucleic acid binding	23993	36	14.27	<1e-30	MF
OG0000422	GO:0090304	nucleic acid metabolic process	2611	13	1.61	0.00014	BP

534

535 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000423	GO:0005515	protein binding	11387	7	1.55	0.0043	MF
	GO:0003674	molecular_function	66532	16	9.03	0.0065	MF
	GO:0110165	cellular anatomical entity	44293	7	7.29	0.00099	CC
OG0000425	GO:0003674	molecular_function	58064	6	4.85	0.0049	MF
OG0000426	GO:0046983	protein dimerization activity	6075	25	1.09	<1e-30	MF
	GO:0005488	binding	51222	28	9.18	0.0077	MF
OG0000430	GO:0003674	molecular_function	71704	39	18.8	0.0029	MF
OG0000433	GO:0046983	protein dimerization activity	6009	32	1.45	<1e-30	MF
OG0000444	GO:0046983	protein dimerization activity	6075	24	1.4	<1e-30	MF
OG0000446	GO:0005488	binding	36717	7	4.79	0.0033	MF
OG0000447	GO:0003674	molecular_function	66532	13	8.7	0.0017	MF
	GO:0003676	nucleic acid binding	21236	5	2.78	0.0085	MF
OG0000449	GO:0016021	integral component of membrane	15812	6	2.14	0.0016	CC
OG0000452	GO:0005488	binding	55053	20	18.77	0.00049	MF
	GO:0005515	protein binding	14711	5	5.01	0.00936	MF
OG0000471	GO:0008270	zinc ion binding	2499	21	0.31	<1e-30	MF
OG0000483	GO:0016021	integral component of membrane	28001	8	2.4	0.0015	CC
OG0000484	GO:0003824	catalytic activity	15658	5	1.73	0.00275	MF
	GO:0110165	cellular anatomical entity	37157	7	9.65	0.00104	CC
OG0000493	GO:0005634	nucleus	8900	40	7.23	<1e-30	CC
OG0000494	GO:0003723	RNA binding	2622	33	0.74	<1e-30	MF
OG0000496	GO:0008270	zinc ion binding	2023	30	0.63	<1e-30	MF
OG0000499	GO:0020037	heme binding	442	19	0.23	<1e-30	MF
	GO:0016020	membrane	36345	39	22.12	<1e-30	CC

536 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000512	GO:0008234	cysteine-type peptidase activity	824	30	0.35	<1e-30	MF
	GO:0003676	nucleic acid binding	25443	10	10.89	0.0054	MF
OG0000523	GO:0005515	protein binding	14711	30	1.92	<1e-30	MF
OG0000525	GO:0003674	molecular_function	64835	36	24.78	0.0036	MF
	GO:0005488	binding	43034	33	16.45	0.0087	MF
OG0000528	GO:0097159	organic cyclic compound binding	30215	17	4.31	0.00013	MF
	GO:1901363	heterocyclic compound binding	30095	16	4.29	0.00438	MF
	GO:1990904	ribonucleoprotein complex	1381	31	0.35	<1e-30	CC
OG0000536	GO:0003676	nucleic acid binding	22734	9	4.09	0.0012	MF
	GO:0005488	binding	48264	27	8.68	0.0018	MF
OG0000541	GO:0003676	nucleic acid binding	22686	39	6.1	<1e-30	MF
OG0000546	GO:0005388	P-type calcium transporter activity	48	23	0.02	<1e-30	MF
OG0000548	GO:0016021	integral component of membrane	24921	6	1.39	0.00015	CC
OG0000561	GO:0003677	DNA binding	4164	22	0.98	<1e-30	MF
	GO:0003674	molecular_function	72442	30	17.12	0.00032	MF
	GO:0110165	cellular anatomical entity	50788	19	13.95	0.0023	CC
OG0000562	GO:0003674	molecular_function	78123	16	8.67	0.00033	MF
OG0000564	GO:0003723	RNA binding	2622	22	0.26	<1e-30	MF
OG0000575	GO:0046983	protein dimerization activity	6236	26	0.87	<1e-30	MF
OG0000577	GO:0016614	oxidoreductase activity, acting on CH-OH...	503	32	0.35	<1e-30	MF
	GO:0050660	flavin adenine dinucleotide binding	472	19	0.33	<1e-30	MF

537

538

539 **Table S8 continuation.** Enrichment analysis of expanded genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000581	GO:0005488	binding	51222	31	12.24	0.00058	MF
	GO:0097159	organic cyclic compound binding	28483	24	6.81	0.00529	MF
OG0000586	GO:0046983	protein dimerization activity	5852	32	1.04	<1e-30	MF
	GO:0003674	molecular_function	67631	36	11.99	0.0096	MF
OG0000590	GO:0110165	cellular anatomical entity	55314	6	5.06	0.00026	CC
OG0000601	GO:0005488	binding	21494	13	11.97	0.00025	MF
OG0000603	GO:0006955	immune response	97	20	0.17	<1e-30	BP
OG0000624	GO:0016021	integral component of membrane	28001	40	14.41	<1e-30	CC
OG0000629	GO:0016765	transferase activity, transferring alkyl...	193	20	0.1	< 1e-30	MF
	GO:0004310	farnesyl-diphosphate farnesyltransferase...	8	8	0	< 1e-30	MF
	GO:0003824	catalytic activity	24890	29	12.73	0.0049	MF
	GO:0016020	membrane	36345	20	19.15	0.0015	CC
OG0000631	GO:0005488	binding	39203	17	5.61	0.00020	MF
	GO:0097159	organic cyclic compound binding	21764	13	3.11	0.00025	MF
OG0000644	GO:0006629	lipid metabolic process	702	26	0.28	<1e-30	BP
OG0000645	GO:0030942	endoplasmic reticulum signal peptide bin...	23	11	0.01	< 1e-30	MF
OG0000653	GO:0004714	transmembrane receptor protein tyrosine ...	15	9	0	<1e-30	MF
OG0000654	GO:0005488	binding	31317	16	10.96	0.00025	MF
	GO:0016020	membrane	16119	6	5.53	0.00068	CC
	GO:0110165	cellular anatomical entity	29181	10	10	0.00184	CC
OG0000661	GO:0004672	protein kinase activity	2357	24	1.21	< 1e-30	MF
OG0000663	GO:0003723	RNA binding	2622	21	0.3	<1e-30	MF
OG0000665	GO:0097159	organic cyclic compound binding	30215	27	5.15	0.00044	MF

541 **Table S8 continuation.** Enrichment analysis of expanded genes.

542

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000666	GO:0005856	cytoskeleton	1716	20	0.51	<1e-30	CC
OG0000667	GO:0003677	DNA binding	4665	31	2.98	<1e-30	MF
OG0000668	GO:0006633	fatty acid biosynthetic process	226	15	0.13	< 1e-30	BP
	GO:0006631	fatty acid metabolic process	244	16	0.14	0.00022	BP
	GO:0016021	integral component of membrane	28001	34	27.22	<1e-30	CC

543

544

545

546

547

548

549

550

551

552

553 **Table S9.** Enrichment analysis of contracted genes.

554

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000005	GO:0043170	macromolecule metabolic process	7729	13	2.4	0.0014	BP
	GO:0044237	cellular metabolic process	8688	13	2.7	0.0025	BP
	GO:0003676	nucleic acid binding	25443	10	9.18	0.00036	MF
	GO:0016021	integral component of membrane	28001	7	11.32	0.0026	CC
OG0000006	GO:0006357	regulation of transcription by RNA polym...	459	39	0.44	<1e-30	BP
	GO:0003690	double-stranded DNA binding	200	20	0.15	< 1e-30	MF
	GO:0005634	nucleus	11058	274	53.15	<1e-30	CC
OG0000007	GO:0005488	binding	42350	33		30 0.00018	MF
	GO:0110165	cellular anatomical entity	42246	6	4.77	0.00392	CC
OG0000014	GO:0005515	protein binding	13234	11	5.85	0.00049	MF
	GO:0046983	protein dimerization activity	6079	5	2.69	0.00283	MF
	GO:0005575	cellular_component	56511	19	23.76	0.00632	CC
	GO:0005634	nucleus	9826	5	4.13	0.00741	CC
	GO:0016020	membrane	32671	9	13.74	0.00803	CC
OG0000017	GO:0003964	RNA-directed DNA polymerase activity	160	15	0.05	< 1e-30	MF
	GO:0003674	molecular_function	78123	25		24 0.00015	MF
	GO:0110165	cellular anatomical entity	54563	6	5.06	0.0010	CC

555

556

557

558 **Table S9 continuation.** Enrichment analysis of contracted genes.

559

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000021	GO:0003676	nucleic acid binding	25443	54	14.82	< 1e-30	MF
	GO:0110165	cellular anatomical entity	61766	20	22.7	0.00028	CC
	GO:0016020	membrane	36345	10	13.36	0.00251	CC
	GO:0005634	nucleus	11058	5	4.06	0.00918	CC
OG0000024	GO:0003676	nucleic acid binding	25443	114	28.14	< 1e-30	MF
	GO:0005488	binding	55053	122	60.88	0.00041	MF
OG0000056	GO:0008234	cysteine-type peptidase activity	577	33	0.52	< 1e-30	MF
	GO:0016021	integral component of membrane	15812	8	6.02	0.00160	CC
OG0000067	GO:0016020	membrane	33578	82	13.19	< 1e-30	CC
OG0000083	GO:0110165	cellular anatomical entity	61766	20	27.24	0.0002	CC
	GO:0043229	intracellular organelle	20891	15	9.21	0.0046	CC
OG0000084	GO:0003676	nucleic acid binding	25443	88	13.01	<1e-30	MF
	GO:0005488	binding	55053	92	28.15	0.0013	MF
	GO:0110165	cellular anatomical entity	61766	5	4.29	0.0040	CC
OG0000086	GO:0003674	molecular_function	71704	56	54.74	0.00021	MF
OG0000088	GO:0003676	nucleic acid binding	22495	51	27	<1e-30	MF
	GO:0003674	molecular_function	71704	56	86.06	0.0041	MF
	GO:0016020	membrane	28700	6	6.21	0.0013	CC
OG0000089	GO:0046983	protein dimerization activity	6236	46	2.55	<1e-30	MF
	GO:0110165	cellular anatomical entity	61766	9	9.08	0.00025	CC
OG0000102	GO:0046983	protein dimerization activity	6075	48	2.57	< 1e-30	MF
	GO:0110165	cellular anatomical entity	54563	6	6.84	0.0031	CC

560 **Table S9 continuation.** Enrichment analysis of contracted genes.

561

562

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000115	GO:0008152	metabolic process	9204	28	11.73	0.0005	BP
	GO:0044238	primary metabolic process	7794	9	9.93	0.0033	BP
	GO:0003676	nucleic acid binding	20350	64	26.52	< 1e-30	MF
	GO:0003674	molecular_function	63780	72	83.11	0.00132	MF
OG0000119	GO:0003676	nucleic acid binding	23993	50	18.86	< 1e-30	MF
	GO:0110165	cellular anatomical entity	55314	9	9.86	0.0011	CC
OG0000121	GO:0005488	binding	46655	10	6.74	0.00051	MF
	GO:0110165	cellular anatomical entity	51078	10	9.51	0.0022	CC
	GO:0016021	integral component of membrane	24237	5	4.51	0.0059	CC
OG0000144	GO:0003676	nucleic acid binding	21236	63	20.82	<1e-30	MF
	GO:0003674	molecular_function	66532	68	65.24	0.0042	MF
	GO:0016021	integral component of membrane	19709	6	4.58	0.00027	CC
OG0000165	GO:0004197	cysteine-type endopeptidase activity	241	63	0.38	< 1e-30	MF
	GO:0003674	molecular_function	84894	84	135.27	0.00096	MF
	GO:0008234	cysteine-type peptidase activity	824	65	1.31	0.00109	MF
OG0000166	GO:0046872	metal ion binding	6890	5	1.41	0.0023	MF
OG0000174	GO:0003676	nucleic acid binding	21042	46	15.77	<1e-30	MF
	GO:0003674	molecular_function	65748	52	49.27	0.0003	MF
	GO:0110165	cellular anatomical entity	44583	8	8.91	0.0017	CC
	GO:0016020	membrane	26404	5	5.28	0.0026	CC

563 **Table S9 continuation.** Enrichment analysis of contracted genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000177	GO:0003674	molecular_function	70920	74	47.15	0.00047	MF
	GO:0003676	nucleic acid binding	22301	33	14.83	0.00505	MF
	GO:0016020	membrane	29171	5	7.1	0.0058	CC
OG0000182	GO:0004252	serine-type endopeptidase activity	982	76	1.05	< 1e-30	MF
OG0000187	GO:0046983	protein dimerization activity	6236	63	3.06	< 1e-30	MF
	GO:0005488	binding	55053	73	27.06	0.0058	MF
OG0000200	GO:0005488	binding	44433	17	18.55	0.0015	MF
	GO:0097159	organic cyclic compound binding	24955	11	10.42	0.0060	MF
	GO:0110165	cellular anatomical entity	44293	7	6.03	0.00098	CC
OG0000209	GO:0110165	cellular anatomical entity	54563	5	6.58	0.0040	CC
OG0000214	GO:0016021	integral component of membrane	28001	72	73.19	< 1e-30	CC
	GO:0016020	membrane	36345	75		95 0.00054	CC
OG0000217	GO:0005488	binding	39203	14	12.62	0.0036	MF
OG0000225	GO:0005488	binding	55053	32	21.82	0.0054	MF
OG0000236	GO:0110165	cellular anatomical entity	54563	6	11.39	0.00026	CC
OG0000258	GO:0110165	cellular anatomical entity	44293	6	6.78	0.0047	CC
OG0000268	GO:0004497	monooxygenase activity	290	46	0.32	< 1e-30	MF
OG0000269	GO:0004190	aspartic-type endopeptidase activity	447	33	0.54	< 1e-30	MF
	GO:0003676	nucleic acid binding	22686	14	27.65	0.00074	MF
	GO:0003674	molecular_function	72951	52	88.93	0.00116	MF
	GO:0110165	cellular anatomical entity	50745	7	8.78	0.00025	CC
OG0000271	GO:0003676	nucleic acid binding	20350	45	10.94	< 1e-30	MF
	GO:0003674	molecular_function	66345	50	35.68	0.00097	MF

564

565 **Table S9 continuation.** Enrichment analysis of contracted genes.

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000273	GO:0003824	catalytic activity	18054	7	2.28	0.00084	MF
OG0000274	GO:0003676	nucleic acid binding	18894	6	2.7	0.0017	MF
OG0000289	GO:0005615	extracellular space	465	69	0.53	<1e-30	CC
OG0000300	GO:0046983	protein dimerization activity	6079	36	1.29	<1e-30	MF
OG0000302	GO:0046983	protein dimerization activity	6009	29	1.23	<1e-30	MF
OG0000346	GO:0003674	molecular_function	70005	7	9.59	0.0047	MF
OG0000376	GO:0003676	nucleic acid binding	22495	6	2.9	0.00018	MF
	GO:0003674	molecular_function	71704	10	9.23	0.00406	MF
	GO:0110165	cellular anatomical entity	48111	5	3.05	0.0042	CC
OG0000388	GO:0003677	DNA binding	4665	26	1.11	< 1e-30	MF
OG0000389	GO:0003674	molecular_function	84894	6	3.7	0.00431	MF
OG0000410	GO:0003824	catalytic activity	14714	41	8.35	< 1e-30	MF
	GO:0016740	transferase activity	4355	7	2.47	0.00057	MF
OG0000413	GO:0046983	protein dimerization activity	5686	22	0.8	<1e-30	MF
OG0000463	GO:0005488	binding	34490	19	9.24	0.0024	MF
OG0000491	GO:0005488	binding	39203	7	5.84	0.00087	MF
	GO:0110165	cellular anatomical entity	38427	26	26.77	0.0019	CC
OG0000503	GO:0003674	molecular_function	66918	18	11.02	0.0098	MF
OG0000509	GO:0005488	binding	39177	25	10.4	0.0039	MF
	GO:0016020	membrane	22090	5	8.25	0.00729	CC
OG0000517	GO:0008234	cysteine-type peptidase activity	576	14	0.11	<1e-30	MF
OG0000524	GO:0005488	binding	44433	8	2.68	0.0067	MF
OG0000538	GO:0140096	catalytic activity, acting on a protein	4710	11	0.93	0.0058	MF

566

567 **Table S9 continuation.** Enrichment analysis of contracted genes.

568

Group	GO.ID	Term	Annotated	Significant	Expected	<i>elim</i>	ontology
OG0000544	GO:0003674	molecular_function	50861	14	9.79	0.00014	MF
	GO:0003824	catalytic activity	13849	7	2.67	0.00070	MF
OG0000547	GO:0016021	integral component of membrane	28001	52	24.01	<1e-30	CC
OG0000566	GO:0005488	binding	47727	16	5.2	0.0062	MF
OG0000589	GO:0110165	cellular anatomical entity	49474	10	11.75	0.00240	CC
OG0000591	GO:0003674	molecular_function	51167	5	2.2	0.0066	MF
	GO:0005634	nucleus	6913	32	7.14	<1e-30	CC
OG0000602	GO:0005575	cellular_component	34591	5	5.07	0.0041	CC
OG0000618	GO:0110165	cellular anatomical entity	46343	11	12.44	0.00053	CC
OG0000657	GO:0003674	molecular_function	53991	24	40.14	0.005	MF
OG0000670	GO:0005488	binding	44433	13	3.8	0.0022	MF

569

570

571

572

573

574

575

576 **Table S10.** Gene detected by Syntenet (Almeida *et al.*, 2023) clustering with
 577 orthogroup (OrthoFinder, Emms and Kelly, 2019) and chemosensory gene. ND = No
 578 defined.
 579

Gene	OR clusters		
	Cluster	Orthogroups	Chemosensory gene group
DPL_DP3025912	389	OG0000359	OR06/OR15/OR16
AGL_AG6002812	4335	ND	ND
AGL_AG6001477	4605	OG0002831	ND
AGL_AG6007490	6067	OG0005707	ND
AGL_AG6027301	9884	OG0000154	OR24
AGL_AG6002816	4336	ND	ND
AGL_AG6007492	6067	OG0005707	ND
AGO_gene-LOC114125477	5228	OG0003362	OR01
AGO_gene-LOC114121820	6067	ND	ND
AGO_Agos_GR1	16430	OG0023414	ND
API_gene-LOC107882832	15982	OG0004751	OR19
API_gene-LOC103311408	15975	OG0011154	OR04
API_gene-LOC100569191	4377	OG0007294	OR02
API_gene-LOC107882864	8139	OG0008724	OR22
API_gene-LOC103309685	12259	ND	ND
API_gene-LOC103308670	12269	OG0000767	OR08b
API_gene-LOC100573389	10730	OG0000183	OR11/OR13/OR14/OR15/OR16
API_gene-LOC103310274	4336	OG0001690	OR08a
API_gene-LOC107883681	8061	OG0000858	OR12
API_gene-LOC115034222	11537	OG0001075	OR08a/OR20/OR21/OR21
API_gene-LOC115034182	6067	OG0004372	ND
API_gene-LOC115034156	6545	OG0010280	OR09
API_gene-LOC107884465	8818	OG0000183	OR11/OR13/OR14/OR15/OR16
API_gene-LOC107883935	6685	OG0000359	OR06/OR15/OR16
API_gene-LOC107884389	4605	OG0002831	ND
API_gene-LOC100165095	5228	OG0003362	OR01
API_gene-LOC107884062	7616	OG0008722	OR05
API_gene-LOC115033244	16719	OG0000154	OR24
API_gene-LOC107882321	9498	OG0001689	OR01/OR03
API_gene-LOC100569439	16751	OG0019062	OR21
API_gene-LOC115033683	15982	OG0008723	OR19
API_gene-LOC107882767	389	OG0000359	OR06/OR15/OR16
API_gene-LOC103309944	6091	OG0001257	OR23
API_gene-LOC107882641	11957	OG0000183	OR11/OR13/OR14/OR15/OR16
API_gene-LOC115034045	7591	OG0006276	OR10
API_gene-LOC100572802	16890	OG0000767	OR08b
API_gene-LOC107883740	17023	OG0000154	OR24
API_gene-LOC107884467	17090	OG0001075	OR08a/OR20/OR21/OR21
API_gene-LOC107884145	17375	OG0001075	OR08a/OR20/OR21/OR21

602 **Table S10 continuation.** Gene detected by Syntenet (Almeida *et al.*, 2023) clustering
 603 with orthogroup (OrthoFinder, Emms and Kelly, 2019) and chemosensory gene.

Gene	OR Clusters		
	Cluster	Orthogroups	Chemosensory gene group
API_gene-LOC100569602	17423	OG0001690	OR08a
API_gene-LOC107885062	16430	OG0000183	OR11/OR13/OR14/OR15/OR16
API_gene-LOC115034265	17507	OG0000359	OR06/OR15/OR16
DNO_gene-LOC107167125	5228	OG0003362	OR01
DNO_gene-LOC107164648	11537	OG0001075	OR08a/OR20/OR21/OR21
DPL_DP3016678	5228	OG0003362	OR01
DPL_DP3027194	11957	OG0000183	OR11/OR13/OR14/OR15/OR16
DPL_DP3005108	10730	OG0000183	OR11/OR13/OR14/OR15/OR16
DPL_DP3003095	12269	OG0000767	OR08b
DPL_DP3003112	12259	OG0008721	OR04
DPL_DP3001802	4377	OG0007294	OR02
DPL_DP3024204	15982	OG0001691	OR18
DPL_DP3024229	15975	OG0011154	OR04
DPL_DP3004842	9498	ND	ND
DPL_DP3025151	7616	OG0008722	OR05
DPL_DP3017402	4605	OG0002831	ND
DPL_DP3025439	17829	OG0000183	OR11/OR13/OR14/OR15/OR16
DPL_DP3026137	6545	OG0013185	OR09
DPL_DP3019081	6091	OG0001257	OR23
DPL_DP3018440	6067	OG0004372	ND
DPL_DP3019609	17860	OG0001578	OR18
DPL_DP3024494	8139	OG0008724	OR22
DPL_DP3024775	8061	OG0000858	OR12
DPL_DP3024194	15982	OG0004751	OR19
DPL_DP3024228	15975	ND	ND
DPL_DP3023139	7591	OG0006276	OR10
DPL_DP3019573	18868	OG0001578	OR18
DPL_DP3007492	16751	OG0001075	OR08a/OR20/OR21/OR21
DPL_DP3024205	19181	OG0001691	OR18
DPL_DP3039696	16719	OG0000154	OR24
ELA_jg6844	11537	ND	ND
ELA_jg11877	15975	OG0011154	OR04
ELA_jg19105	19552	OG0011826	OR07
MPE_g16576	15975	OG0011154	OR04
MPE_g26049	4336	OG0001690	OR08a
MPE_g16623	14204	OG0000154	OR24
PNI_g23800	6067	OG0004372	ND
PNI_g19423	17059	ND	ND
RMA_gene-LOC113561217	6067	ND	ND

604 **Table S10 continuation.** Gene detected by Syntenet (Almeida *et al.*, 2023) clustering
 605 with orthogroup (OrthoFinder, Emms and Kelly, 2019) and chemosensory gene.

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

Gene	OR Clusters		
	Cluster	Orthogroups	Chemosensory gene group
DPL_DP3052799	389	OG0000359	OR06/OR15/OR16
DPL_DP3052805	389	OG0000359	OR06/OR15/OR16
DPL_DP3024766	8061	OG0000858	OR12
DPL_DP3024288	14205	ND	ND
ELA_jg6846	14606	ND	ND
RPA_g1216	6067	OG0004372	ND
SMI_SMIS_genome.masked024709	4336	ND	ND
SMI_SMIS_genome.masked013819	6067	OG0004372	ND
SMI_SMIS_genome.masked014025	6067	OG0005707	ND
DPL_DP3024312	16715	OG0000154	OR24
DPL_DP3024324	16717	OG0000154	OR24
DPL_DP3024298	16718	OG0000154	OR24
DPL_DP3025904	389	OG0000359	OR06/OR15/OR16
DPL_DP3019084	16804	OG0001257	OR23
DPL_DP3006992	17538	OG0000183	OR11/OR13/OR14/OR15/OR16
MPE_g8463	18318	OG0000154	OR24
SMI_SMIS_genome.masked020211	14204	OG0016760	ND
RPA_g1213	6067	OG0004372	ND

629 **Table S10 continuation.** Gene detected by Syntenet (Almeida *et al.*, 2023) clustering
 630 with orthogroup (OrthoFinder, Emms and Kelly, 2019) and chemosensory gene.

Gene	GR Clusters		
	Cluster	Orthogroup	Chemosensory gene group
AGL_AG6001192	4860	OG0021379	ND
AGL_AG6004420	13256	ND	ND
AGL_AG6011285	14067	OG0010875	ND
AGO_gene-LOC114122052	11241	ND	ND
AGO_gene-LOC114122707	9723	ND	ND
AGO_gene-LOC114126236	9172	ND	ND
API_gene-LOC100162516	7856	OG0000952	GR01/GR07
API_gene-LOC100166624	7856	OG0000952	GR01/GR07
API_gene-LOC100570455	16180	OG0008718	GR02
API_gene-LOC100570505	7856	OG0000952	GR01/GR07
API_gene-LOC100571929	9723	OG0004749	ND
API_gene-LOC100573741	11241	OG0001969	GR08/GR10
API_gene-LOC100574450	14721	OG0010277	GR11
API_gene-LOC100574455	5471	OG0005411	GR01
API_gene-LOC100575410	9509	OG0001418	GR07
API_gene-LOC100575495	14067	OG0010875	ND
API_gene-LOC103307698	5348	OG0003048	GR09
API_gene-LOC103309266	11241	OG0015402	ND
API_gene-LOC107882751	9172	OG0005412	GR01
API_gene-LOC107882782	7894	OG0000694	GR05
API_gene-LOC107883586	11981	OG0000282	GR04
API_gene-LOC107883663	16179	OG0005442	ND
API_gene-LOC107884112	17135	OG0010361	GR06
API_gene-LOC107884336	6040	OG0002789	GR07
API_gene-LOC107884492	9509	OG0006275	GR07
API_gene-LOC107884679	10569	OG0000282	GR04
API_gene-LOC107885062	16430	OG0000183	OR11/OR13/OR14/OR15/OR16
API_gene-LOC115033181	10569	OG0000282	GR04
API_gene-LOC115033906	7856	OG0015319	ND
API_gene-LOC115034294	13822	OG0010277	GR11
DPL_DP3003222	8608	OG0010276	ND
DPL_DP3006121	13256	OG0011058	ND
DPL_DP3013752	18165	ND	ND
DPL_DP3020709	7247	OG0000745	ND
DPL_DP3020716	7247	OG0000745	ND
DPL_DP3024538	16179	OG0005442	ND
DPL_DP3026652	17838	OG0000282	GR04

631

632 **Table S10 continuation.** Gene detected by Syntenet (Almeida *et al.*, 2023) clustering
 633 with orthogroup (OrthoFinder, Emms and Kelly, 2019) and chemosensory gene.

634

Gene	GR Clusters		
	Cluster	Orthogroup	Chemosensory gene group
ELA_jg16045	13256	OG0011058	ND
ELA_jg17687	8608	OG0010276	ND
ELA_jg20410	14067	OG0010875	ND
ELA_jg20749	13822	OG0003048	GR09
ELA_jg20906	16180	OG0008718	GR02
MCE_Mca06485	11241	OG0003048	GR09
MCE_Mca07992	14067	OG0010875	ND
MCE_Mca09929	16179	OG0005442	ND
MCE_Mca09930	16180	OG0008718	GR02
MPE_g12505	17838	OG0000282	GR04
MPE_g14270	9509	OG0006275	GR07
MPE_g19340	16179	OG0005442	ND
PNI_g21031	14129	OG0010279	GR06
PNI_g23951	9723	OG0004749	ND
RMA_gene-LOC113550714	7247	OG0000745	ND
RMA_gene-LOC113554567	14067	OG0010875	ND
RMA_gene-LOC113555031	15022	OG0000282	GR04
RMA_gene-LOC113558094	9723	ND	ND
RMA_gene-LOC113558297	7218	OG0001074	GR03
RPA_g18293	4860	ND	ND
RPA_g25374	9723	OG0004749	ND
RPA_g25376	9723	OG0004749	ND
SMI_SMIS_genome.masked001507	13256	OG0011058	ND
SMI_SMIS_genome.masked001508	13256	ND	ND
SMI_SMIS_genome.masked012442	7218	ND	ND
SMI_SMIS_genome.masked019221	7856	ND	ND
SMI_SMIS_genome.masked019222	15566	OG0015319	ND
SMI_SMIS_genome.masked019804	6040	OG0002789	GR07
SMI_SMIS_genome.masked023058	3907	ND	ND
SMI_Smis_OR33	16430	OG0000183	OR11/OR13/OR14/OR15/OR16
DPL_DP3037636	7856	ND	ND
DPL_DP3040769	14067	OG0010875	ND

635

636

637

638

639

640

641

642

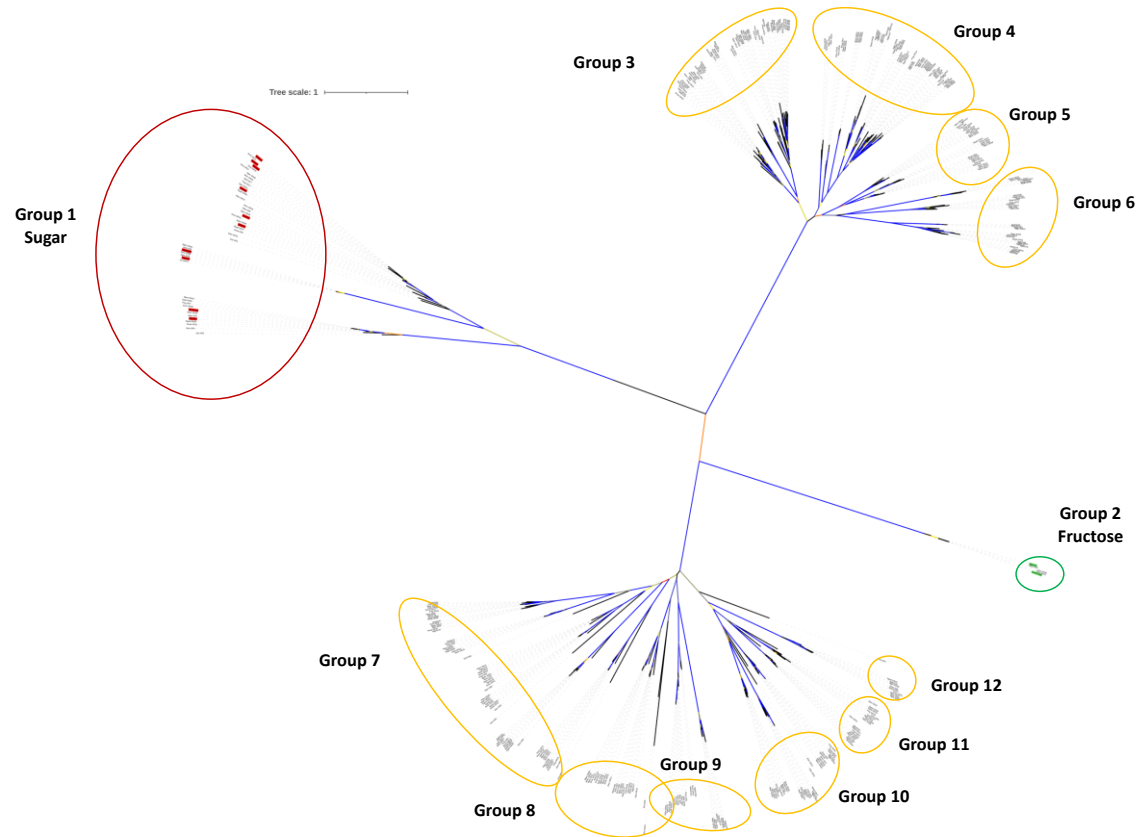
643

644

645

646

647



648 **Figure S7.** Twelve groups of the candidate gustatory receptors (GR) genes were detected using a phylogenetic tree. Sequences were
649 aligned using MAFFT, and the phylogenetic tree was reconstructed using IQTree software. The sugar and fructose gustatory receptors
650 genes were represented in red and green circles, respectively. Bootstrap values were represented in the branches from the lowest
651 80% in red to the highest 100% in blue.

652

653

654

655

656

657

658

659

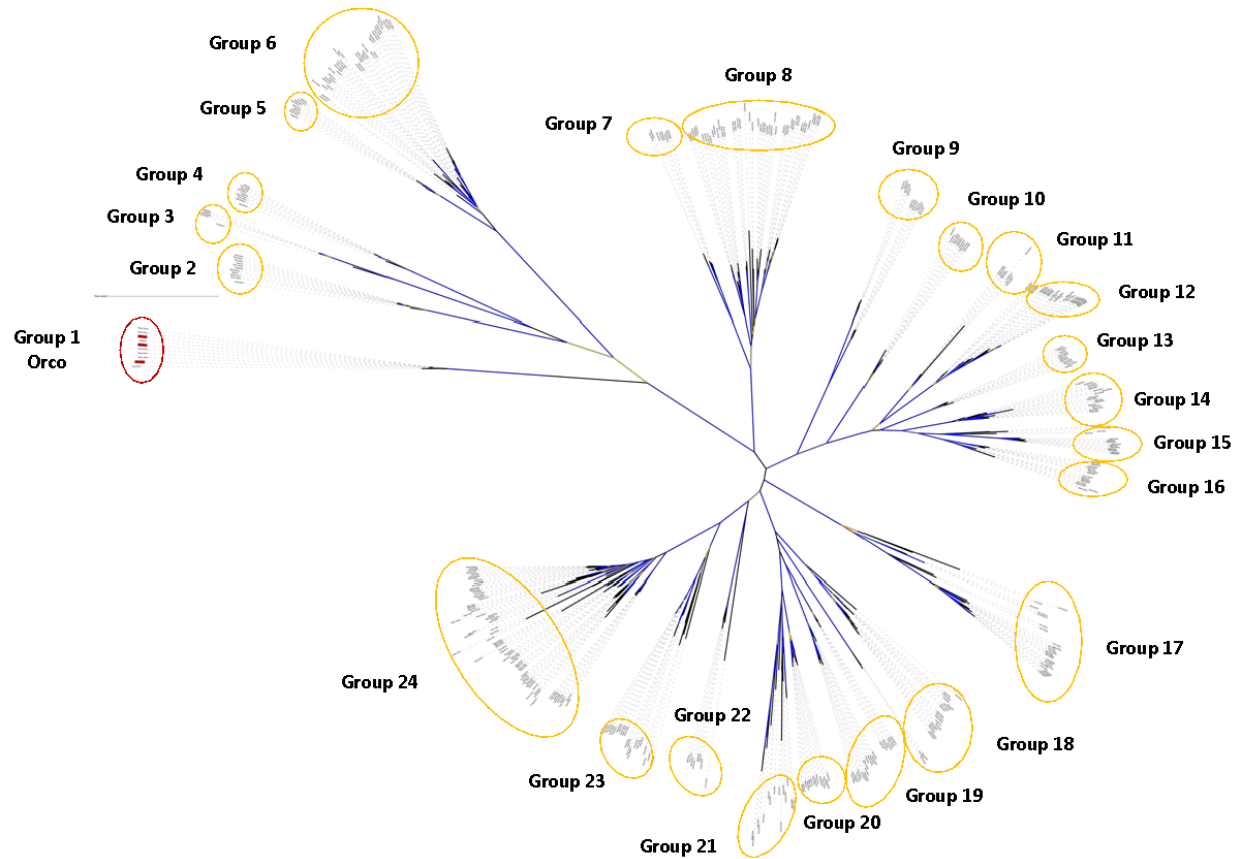
660

661

662

663

664



665

666

667

668

Figure S8. Twenty-four groups of the candidate olfactory receptors (OR) genes were detected using a phylogenetic tree. Sequences were aligned using MAFFT, and the phylogenetic tree was reconstructed using IQTree. The Orco olfactory receptors are represented in red color. Bootstrap values were represented in the branches from the lowest 80% in red to the highest 100% in blue.

669 **Table S11.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) branch model summary statistics inferred
 670 from PAML v4 (Yang, 2007).

671

	GR1 Sugar	GR2 Fructose	GR3	GR4	GR5	GR6	GR7	GR8	GR9	GR10	GR11	GR12
N	44	5	56	57	25	32	67	24	22	32	19	10
S	354	285	234	408	519	456	441	822	441	639	858	789
ω	0.56463	0.88726	1.02950	0.72983	0.95776	0.87368	0.75979	0.79087	1.28869	0.93441	0.48488	0.56346
H_oInL	-5231.688326	-1216.789547	-5789.109381	-9980.164042	-5478.870866	-6048.019804	-12088.855011	-9355.605423	-4445.296008	-9732.432759	-8102.421100	-4029.935943
H_aInL	-5185.375442	-1213.859064	-5738.300335	-9919.114544	-5457.974317	-6023.154680	-12017.378990	-9305.779103	-4418.048900	-9701.739763	-8074.971065	-4023.319379
2ΔL	92.62577	5.860966	101.62	122.1	41.793	49.73	142.95	99.653	54.49	61.386	54.9	13.233
K	43	4	55	56	24	31	66	23	21	31	18	9
alpha	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
p	1.702e-05	0.2098	0.0001335	8.08e-07	0.01361	0.01783	1.321e-07	1.616e-11	8.374e-05	0.0009238	1.327e-05	0.1523
χ^2	59.304	9.4877	73.311	74.468	36.415	44.985	85.965	35.172	32.671	44.985	28.869	16.919

672 **N=** number of genes; **S =** number of sites; **$\omega=$** dN/dS value; **H_o (InL) :** One ratio hypothesis likelihood score; **H_a (InL) :** Free ratio hypothesis likelihood score; **K :**Degrees of freedom; **alpha :** 0.05 ; **p-value :** <0.05 ; **χ^2 :** chi-square
 673 distribution; GR = Gustatory receptor group

674

675

676

677

678

679 **Table S11 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) branch model summary
 680 statistics inferred from PAML v4 (Yang, 2007).

681

682

	OR1 Orco	OR2	OR3	OR4	OR5	OR6	OR7	OR8a	OR8b	OR9	OR10	OR11	OR12
N	10	11	5	9	9	37	8	15	31	12	10	8	25
S	270	1128	1104	795	795	312	1002	444	60	939	1011	579	468
ω	0.94407	0.67866	0.43454	1.24689	1.50124	0.84971	0.68829	1.47700	1.87058	1.73311	1.03558	0.99001	1.20879
H_0InL	-	-	-	-	-	-	-	-	-	-	-	-	-
	1268.682720	4855.073809	3317.411588	3848.094699	3564.354753	3407.882779	-4468.344612	-3627.976230	-856.474042	5318.031193	-3850.311741	2047.397742	-3586.931783
H_aInL	-	-	-	-	-	-	-	-	-	-	-	-	-
	1259.470606	4839.687812	3312.508079	3842.027866	3557.624712	3362.358362	-4459.555754	-3616.782859	-834.696518	5313.399292	-3844.135355	2043.921807	-3569.650301
$2\Delta L$	18.424	30.772	9.807	12.134	13.46	91.049	17.578	22.387	43.555	9.2638	12.353	6.9519	34.563
K	9	10	4	8	8	36	7	14	30	11	9	7	24
alpha	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
p	0.03056	0.0006398	0.04381	0.1453	0.09697	1.156e-06	0.01403	0.07101	0.05231	0.5976	0.1941	0.4339	0.07515
χ^2	16.919	18.307	9.4877	15.507	15.507	50.998	14.067	23.685	43.773	19.675	16.919	14.067	36.415

683

684 **N**= number of genes; **S** = number of sites; **ω** = dN/dS value; **H_0 (InL)** : One ratio hypothesis likelihood score; **H_a (InL)** : Free ratio hypothesis likelihood score; **K** :Degrees of freedom; **alpha** : 0.05 ; **p-value** : <0.05 ; **χ^2** : chi-square distribution; OR = Olfactory receptor group

685

686

687

688

689 **Table S11 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) branch model summary
 690 statistics inferred from PAML v4 (Yang, 2007).

691

	OR13	OR14	OR15	OR16	OR17	OR18	OR19	OR20	OR21	OR22	OR23
N	11	17	14	15	25	21	21	14	15	8	22
S	885	306	678	378	387	702	1062	1086	195	618	300
ω	0.44286	1.29363	0.84264	0.77929	0.50600	0.92750	0.45569	0.76303	0.87819	1.11350	1.35136
H_oInL	-3745.413397	-2194.749421	-5204.634667	-2238.207750	-4932.915224	-8533.306542	-8986.854202	-6162.496007	-2123.429093	-3328.270610	-2677.570631
H_aInL	-3704.800841	-2179.939307	-5187.516447	-2227.199444	-4898.485935	-8510.967430	-8939.307333	-6147.827618	-2109.404275	-3320.568051	-2656.982334
2ΔL	81.225	29.62	34.236	22.017	68.859	44.678	95.094	29.337	28.05	15.405	41.177
K	10	16	13	14	24	20	20	13	14	7	21
alpha	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
p	2.887e-13	0.02007	0.001108	0.07827	3.251e-06	0.00122	9.411e-12	0.005862	0.01401	0.03114	0.005332
χ^2	18.307	26.296	22.362	23.685	36.415	31.41	31.41	22.362	23.685	14.067	32.671

692

693 **N=** number of genes; **S =** number of sites; **ω =** *dN/dS* value; **H_o (InL) :** One ratio hypothesis likelihood score; **H_a (InL) :** Free ratio hypothesis likelihood score; **K :**Degrees of freedom; **alpha :** 0.05 ; **p-value :** <0.05 ; **χ^2 :** chi-square
 694 distribution; OR = Olfactory receptor group

695

696

697

698 **Table S12.** Chemosensory genes (gustatory receptors GR and olfactory receptors OR) were
 699 suggested under positive ($\omega>0$), purifying ($\omega<0$), and neutral ($\omega=0$) selection based on the
 700 branch model analysis PAML v4 (Yang, 2007).

701

Genes	Numer of genes	Sites	Free Omega	Number of sequences with $\omega>0$	Number of sequences with $\omega<0$	Number of sequences with $\omega=0$	Saturated
GR01	44	354	Yes	1	29	14	0
GR02	5	285	No	0	4	1	0
GR03	56	234	Yes	15	22	19	0
GR04	57	408	Yes	12	34	11	0
GR05	25	519	Yes	5	14	6	0
GR06	32	456	Yes	10	16	6	0
GR07	67	441	Yes	14	40	13	0
GR08	24	822	Yes	5	17	2	0
GR09	22	441	Yes	10	6	6	0
GR10	32	639	Yes	12	17	3	0
GR11	19	858	Yes	0	19	0	0
GR12	10	789	No	1	9	0	0
Total	393	6246		85	227	81	0

702

703

704

705

706

707

708

709

710

711

712

713

714

715 **Table S12 continuation.** Chemosensory genes (gustatory receptors GR and olfactory
 716 receptors OR) were suggested under positive ($\omega>0$), purifying ($\omega<0$), and neutral ($\omega=0$)
 717 selection based on the branch model analysis PAML v4 (Yang, 2007).

718

719

Genes	Numer of genes	Sites	Free Omega	Number of sequences with $\omega>0$	Number of sequences with $\omega<0$	Number of sequences with $\omega=0$	Saturated
OR01 (Orco)	10	270	Yes	1	4	3	0
OR02	11	1128	Yes	2	8	1	0
OR03	5	1104	Yes	1	4	0	0
<u>OR04</u>	<u>9</u>	<u>795</u>	<u>No</u>	<u>7</u>	<u>0</u>	<u>2</u>	0
OR05	9	795	Yes	2	3	4	0
OR06	37	312	Yes	7	19	11	0
OR07	8	1002	Yes	1	4	3	0
<u>OR08a</u>	<u>15</u>	<u>444</u>	<u>No</u>	<u>6</u>	<u>3</u>	<u>5</u>	1
<u>OR08b</u>	<u>31</u>	<u>60</u>	<u>No</u>	1	<u>3</u>	<u>27</u>	0
<u>OR09</u>	<u>12</u>	<u>939</u>	<u>No</u>	<u>7</u>	1	<u>4</u>	0
<u>OR10</u>	<u>10</u>	<u>1011</u>	<u>No</u>	<u>2</u>	<u>5</u>	<u>3</u>	0
<u>OR11</u>	<u>8</u>	<u>579</u>	<u>No</u>	<u>2</u>	<u>4</u>	<u>2</u>	0
<u>OR12</u>	<u>25</u>	<u>468</u>	<u>No</u>	<u>8</u>	<u>11</u>	<u>6</u>	0
OR13	11	885	Yes	1	9	1	0
OR14	17	306	Yes	4	10	3	0
OR15	14	678	Yes	4	10	0	0
<u>OR16</u>	<u>15</u>	<u>378</u>	<u>No</u>	<u>6</u>	<u>6</u>	<u>2</u>	1
OR17	25	387	Yes	3	15	7	0
OR18	21	702	Yes	10	10	1	0
OR19	21	1062	Yes	2	18	1	0
OR20	14	1086	Yes	2	12	0	0
OR21	15	195	Yes	4	8	3	0
OR22	8	618	Yes	3	3	2	0
OR23	22	300	Yes	9	2	11	0
Total	373	15504		95	172	102	2

720

721

722 **Table S13.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) site model summary statistics inferred from
 723 PAML v4 (Yang, 2007).

724

Model	GR1 Sugar	GR2 Fructose	GR 3	GR4	GR5	GR6	GR7	GR8	GR9	GR10	GR11	GR12	
M0	N	44	5	56	57	25	32	67	24	22	32	19	10
	S	354	285	234	408	519	456	441	822	441	639	858	789
	ω	0.56979	0.88676	1.03516	0.76955	0.94636	0.89746	0.77104	0.80583	1.29361	0.93989	0.48382	0.57127
	InL	-5231.825340	-1216.789767	-5789.152113	-9998.205966	-5479.083036	-6049.200897	-12089.821784	-9356.298793	-4445.435595	-9732.621983	-8102.433935	-4030.348531
M1a	p0	0.83458	0.30857	0.83073	0.86967	0.76096	0.80609	0.86730	0.69101	0.61788	0.75584	0.82457	0.69644
	p1	0.16542	0.69143	0.16927	0.13033	0.23904	0.19391	0.13270	0.30899	0.38212	0.24416	0.17543	0.30356
	InL	-5077.671870	-1212.480402	-5490.213977	-9402.039709	-5353.305732	-5767.175345	-11420.967944	-9101.153882	-4366.772067	-9296.978281	-7794.902158	-3961.899495

725

726

727

728

729

730

731

732 **Table S13 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) site model summary statistics
 733 inferred from PAML v4 (Yang, 2007).

734	Model	GR1 Sugar	GR2 Fructose	GR 3	GR4	GR5	GR6	GR7	GR8	GR9	GR10	GR11	GR12	
735	M2a	p_0	0.81467	0.58851	0.76326	0.82639	0.65602	0.75180	0.82921	0.62212	0.58002	0.69137	0.81508	0.76591
		p_1	0.18122	0.00000	0.21946	0.15170	0.30452	0.20453	0.14031	0.33966	0.27455	0.25551	0.16766	0.15610
736		p_2	0.00410	0.41149	0.01729	0.02191	0.03946	0.04366	0.03048	0.03822	0.14543	0.05312	0.01726	0.07799
		ω_2	3.97984	1.51071	3.76954	2.85535	3.96942	2.87054	2.34407	3.04424	2.82709	2.67130	2.65679	2.12447
737		InL	5059.921483	1211.381107	5397.045087	9304.423044	5289.395383	5713.611109	11342.326713	9053.610777	4315.948000	9222.323267	7772.779594	3953.982877
738	M3	p_0	0.58332	0.58851	0.76552	0.80958	0.73032	0.78935	0.76004	0.53892	0.62849	0.68595	0.76264	0.80718
739		p_1	0.35161	0.13611	0.21827	0.16392	0.23582	0.19702	0.17272	0.36640	0.26443	0.25517	0.19963	0.18940
		p_2	0.06507	0.27538	0.01622	0.02650	0.03386	0.01363	0.06724	0.09467	0.10708	0.05887	0.03773	0.00341
740		ω_0	0.04657	0.29695	0.08337	0.06080	0.26344	0.09879	0.04973	0.02481	0.08967	0.03038	0.04025	0.15817
741		ω_1	0.45267	1.51071	1.13408	0.79733	1.21337	1.38720	0.58398	0.64228	1.30665	0.94832	0.70367	1.52521
742	ω_2	1.72582	1.51072	4.16281	2.50663	4.27373	4.26718	1.78145	2.16544	3.07725	2.56541	2.13370	4.41213	
	InL	5052.892647	1211.381107	5396.002334	9300.606905	5289.011093	5710.513088	11335.539722	9049.122334	4315.713162	9222.209445	7770.677152	3953.943748	

743

744

745

746

747

748 **Table S13 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) site model summary statistics
 749 inferred from PAML v4 (Yang, 2007).

750

751

Model	GR1 Sugar	GR2 Fructose	GR 3	GR4	GR5	GR6	GR7	GR8	GR9	GR10	GR11	GR12	
M7	<i>p</i>	0.17455	0.03469	0.05396	0.04292	0.14846	0.01359	0.04789	0.09230	0.01421	0.01600	0.07647	0.11235
	<i>q</i>	0.56922	0.01280	0.27935	0.26699	0.28807	0.04376	0.29159	0.19519	0.01886	0.05103	0.29329	0.18386
	<i>lnL</i>	5077.945793	1212.499290	5480.009432	9379.699699	5365.565151	5773.318298	-11401.784294	9100.108902	4367.195585	9293.315424	7804.778381	3966.461479
M8	<i>p₀</i>	0.95375	0.59059	0.98039	0.97635	0.95253	0.92994	0.95670	0.94099	0.84540	0.93706	0.96119	0.86209
	<i>p₁</i>	0.04625	0.40941	0.01961	0.02365	0.04747	0.07006	0.04330	0.05901	0.15460	0.06294	0.03881	0.13791
	<i>p</i>	0.43013	42.38437	0.01815	0.10074	0.46429	0.15777	0.12647	0.18259	0.12182	0.07722	0.17162	1.22942
	<i>q</i>	1.49525	99.00000	0.05848	0.50118	0.60192	0.53504	0.63448	0.38289	0.22543	0.22473	0.79586	4.67910
	<i>ω</i>	1.89569	1.51316	3.58829	2.50414	3.69320	2.51872	1.97278	2.47927	2.78821	2.42426	2.09829	1.83933
	<i>lnL</i>	5053.994057	1211.382858	5394.753409	9290.460615	5290.583339	5716.276926	11327.505439	9048.607034	4316.436398	9221.616983	7771.030353	3954.014079

752

753

754

755

756

757 **Table S13 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) site model summary statistics
 758 inferred from PAML v4 (Yang, 2007).

759

	OR1 Orco	OR2	OR3	OR4	OR5	OR6	OR7	OR8a	OR8b	OR9	OR10	OR11
<i>N</i>	10	11	5	9	9	37	8	15	31	12	10	8
<i>S</i>	270	1128	1104	795	795	312	1002	444	60	939	1011	579
<i>M0</i> ω	0.71014	0.67310	0.47414	1.24881	1.39432	0.85984	0.75226	1.50944	1.87082	1.81900	0.94383	0.98620
<i>InL</i>	-	-	-	-	-	-	-	-	-	-	-	-
	1277.419980	4855.183909	3324.042385	3848.105608	3568.098925	3408.209158	4479.693944	3631.526679	856.474056	5319.972747	3857.677939	2047.402217
<i>M1a</i> p_0	0.36814	0.55993	0.71671	0.34264	0.47116	0.76251	0.60290	0.57692	0.72571	0.52422	0.44514	0.39714
p_1	0.63186	0.44007	0.28329	0.65736	0.52884	0.23749	0.39710	0.42308	0.27429	0.47578	0.55486	0.60286
<i>InL</i>	-	-	-	-	-	-3296.34482	-	-	-	-	-	-
	1259.699478	4797.822208	3308.891909	3832.394693	3543.684662		4396.215633	3570.720139	827.407331	5267.011477	3823.483550	2037.472383

760

761

762

763

764

765

766

767 **Table S13 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) site model summary statistics
 768 inferred from PAML v4 (Yang, 2007).

769

770

	OR1 Orco	OR2	OR3	OR4	OR5	OR6	OR7	OR8a	OR8b	OR9	OR10	OR11	
771	<i>p₀</i>	0.21281	0.53032	0.71671	0.67958	0.30517	0.71475	0.65679	0.53348	0.50886	0.27401	0.38696	0.63895
	<i>p₁</i>	0.74794	0.46054	0.15964	0.12322	0.61663	0.24850	0.26918	0.25713	0.41403	0.60330	0.54882	0.18189
772	<i>M2a p₂</i>	0.03925	0.00914	0.12365	0.19720	0.07820	0.03676	0.07403	0.20939	0.07711	0.12270	0.06423	0.17916
	<i>ω₂</i>	7.15980	5.78168	1.00000	3.11829	6.07725	2.77295	2.72780	2.98276	6.40163	5.40352	4.57668	2.77583
773	<i>InL</i>	1251.776345	4781.144351	3308.891909	3801.445999	3497.405393	3276.383988	4382.104041	3522.928259	786.311979	5179.266618	3799.778882	2031.484327
774	<i>p₀</i>	0.32170	0.71351	0.31804	0.05548	0.72186	0.74631	0.74238	0.59648	0.51417	0.41306	0.65438	0.19889
775	<i>p₁</i>	0.64604	0.28469	0.67990	0.74503	0.27074	0.23425	0.24481	0.29658	0.40955	0.48803	0.32655	0.62448
	<i>p₂</i>	0.03227	0.00180	0.00206	0.19949	0.00740	0.01944	0.01282	0.10694	0.07628	0.09891	0.01907	0.17663
776	<i>M3 ω₀</i>	0.25649	0.12934	0.00000	0.00000	0.31501	0.08512	0.13582	0.10659	0.00000	0.11346	0.19958	0.00000
	<i>ω₁</i>	1.25267	1.51353	0.60434	0.58481	3.03382	1.27091	1.55645	1.63950	1.04134	1.36181	1.80986	0.59144
777	<i>ω₂</i>	8.23044	10.27914	5.54959	3.11919	14.95929	3.40857	4.21949	3.66459	6.52613	6.01030	6.72778	2.81656
778	<i>InL</i>	1251.614549	4779.219529	3307.756934	3801.408614	3495.220514	3275.335808	4382.009169	3522.071847	786.305067	5179.035727	3799.432971	2031.469449

779

780

781

782

783 **Table S13 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) site model summary statistics
 784 inferred from PAML v4 (Yang, 2007).

785

	OR1 Orco	OR2	OR3	OR4	OR5	OR6	OR7	OR8a	OR8b	OR9	OR10	OR11	
M7	<i>p</i>	0.56952	0.00534	0.53125	0.03523	4.36286	0.06430	0.01469	0.01169	0.01581	0.01886	0.00751	0.00726
	<i>q</i>	0.22355	0.00663	0.72315	0.01998	0.00500	0.16543	0.01920	0.00500	0.09509	0.01688	0.00500	0.00500
	<i>InL</i>	-	-4798.331056	-3308.277304	-3833.774262	-3569.585274	-3303.021239	-4396.427598	-3589.963693	-827.608477	-5267.344755	-3823.955115	-2037.473281
		8											
	<i>p₀</i>	0.95789	0.99135	0.99906	0.80324	0.92481	0.94742	0.90001	0.77658	0.92524	0.87820	0.93845	0.82429
	<i>p₁</i>	0.04211	0.00865	0.00094	0.19676	0.07519	0.05258	0.09999	0.22342	0.07476	0.12180	0.06155	0.17571
M8	<i>p</i>	0.45550	0.00516	0.59207	3.49022	0.01159	0.13305	0.27595	0.15528	0.01774	0.01159	0.00739	0.97790
	<i>q</i>	0.11128	0.00502	0.82021	2.84556	0.00500	0.32512	0.60179	0.28619	0.02063	0.00500	0.00500	1.17869
	<i>ω</i>	6.96738	6.08225	6.10592	3.12781	6.31042	2.55244	2.51752	2.94391	6.46621	5.46641	4.69363	2.80853
	<i>InL</i>	-	-4781.370179	-3308.148161	-3801.432365	-3497.460429	-3279.169203	-4382.169578	-3523.414754	-786.377213	-5179.281417	-3799.798015	-2031.476879
		2											

786

787

788

789

790

791 **Table S13 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) site model summary statistics
 792 inferred from PAML v4 (Yang, 2007).

793

794

		OR12	OR13	OR14	OR15	OR16	OR17	OR18	OR19	OR20	OR21	OR22	OR23
M0	N	25	11	17	14	15	25	21	21	14	15	8	22
	S	468	885	306	678	378	387	702	1062	1086	195	618	300
	ω	1.16038	0.44317	1.29761	0.93656	0.84268	0.53613	1.04444	0.47798	0.75217	0.92974	1.13172	1.29806
	<i>lnL</i>	-3588.139836	-3745.413691	-2194.753612	-5214.486356	-2242.826138	-4940.105104	-8552.735563	-8993.558702	-6162.761013	-2125.005660	-3330.081615	-2678.888227
M1a	p_0	0.68868	0.79216	0.59661	0.70533	0.69271	0.83371	0.75104	0.81679	0.67477	0.73123	0.38003	0.67300
	p_1	0.31132	0.20784	0.40339	0.29467	0.30729	0.16629	0.24896	0.18321	0.32523	0.26877	0.61997	0.32700
	<i>lnL</i>	-3486.586634	-3681.069046	-2158.354702	-5063.053386	-2186.389432	-4711.510404	-8191.950929	-8678.778617	-6020.326260	-2036.359161	-3304.977694	-2617.608997

795

796

797

798

799

800

801

802 **Table S13 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) site model summary statistics
 803 inferred from PAML v4 (Yang, 2007).

804

805

806

807

808

809

810

811

812

813

814

815

816

817

	OR12	OR13	OR14	OR15	OR16	OR17	OR18	OR19	OR20	OR21	OR22	OR23
M2a p_0	0.59234	0.83799	0.42401	0.64662	0.59354	0.82025	0.64308	0.80504	0.61277	0.64163	0.31707	0.53927
p_1	0.31969	0.14657	0.51653	0.29392	0.38315	0.17491	0.30966	0.18309	0.36564	0.33663	0.61026	0.36370
p_2	0.08797	0.01544	0.05945	0.05946	0.02331	0.00484	0.04725	0.01187	0.02159	0.02174	0.07267	0.09703
ω_2	3.77467	4.40582	4.43938	3.07580	5.99272	3.49845	3.49847	2.79929	4.95223	3.60716	3.85225	3.97549
<i>InL</i>	3438.453620	3662.868100	2131.706907	5031.428120	2156.215036	4701.791333	8097.582281	8663.562640	5963.486385	2021.667637	3284.332897	2574.028379
M3 p_0	0.67486	0.58639	0.58207	0.65284	0.52051	0.64560	0.63140	0.73960	0.73515	0.64471	0.37781	0.60685
p_1	0.26661	0.39277	0.39502	0.29005	0.45173	0.27770	0.31566	0.23038	0.25744	0.33664	0.57315	0.31410
p_2	0.05853	0.02084	0.02291	0.05711	0.02776	0.07670	0.05294	0.03002	0.00741	0.01865	0.04904	0.07905
ω_0	0.04735	0.04683	0.10372	0.10828	0.00000	0.00435	0.04540	0.03742	0.07173	0.00000	0.04229	0.13042
ω_1	1.45850	0.54834	1.61123	1.02520	0.82450	0.39270	0.91872	0.67852	1.48937	1.07150	1.17034	1.28588
ω_2	4.41467	3.99716	6.64091	3.11859	5.48638	1.38536	3.28300	2.12257	7.27234	3.96416	4.43243	4.34745
<i>InL</i>	3437.771521	3662.608215	2130.789706	5031.424098	2155.939738	4695.518331	8097.321728	8659.143002	5961.340929	2021.568078	3284.228800	2573.756787

818 **Table S13 continuation.** Chemosensory genes (Gustatory receptors - GR and Olfactory receptors - OR) site model summary statistics
 819 inferred from PAML v4 (Yang, 2007).

820

	OR12	OR13	OR14	OR15	OR16	OR17	OR18	OR19	OR20	OR21	OR22	OR23	
M7	<i>p</i>	0.00945	0.06887	0.00856	0.09476	0.01392	0.09327	0.06380	0.09623	0.01291	0.01452	0.02078	0.01675
	<i>q</i>	0.01886	0.17861	0.00576	0.17918	0.02706	0.42014	0.19649	0.36116	0.02602	0.02855	0.01255	0.03282
	<i>lnL</i>	-3486.779859	-3684.323225	-2164.038648	-5071.873297	-2186.708961	-4703.931918	-8190.929871	-8685.315986	-6020.805255	-2034.553460	-3305.056752	-2621.680496
	<i>p₀</i>	0.90256	0.98037	0.93716	0.91468	0.97500	0.98193	0.94664	0.97281	0.97938	0.98072	0.91736	0.89775
	<i>p₁</i>	0.09744	0.01963	0.06284	0.08532	0.02500	0.01807	0.05336	0.02719	0.02062	0.01928	0.08264	0.10225
M8	<i>p</i>	0.01433	0.43423	0.00597	0.29272	0.09010	0.13751	0.13389	0.17680	0.00500	0.01672	0.03341	0.18363
	<i>q</i>	0.02810	1.28183	0.00500	0.58188	0.13759	0.61739	0.29295	0.73183	0.00758	0.03256	0.01869	0.22021
	<i>ω</i>	3.58020	4.07223	4.23499	2.71380	5.73082	2.05538	3.11261	2.15466	5.15913	3.52442	3.63801	3.93879
	<i>lnL</i>	-3438.692997	-3662.634207	-2131.892309	-5031.964759	-2156.051260	-4693.116033	-8097.050353	-8659.506895	-5963.772860	-2021.029561	-3284.481367	-2574.659541

821

822

823

824

825

826

827

828 **Table S14.** Chemosensory genes (gustatory genes-GR and olfactory genes-OR) likelihood-rate test summary of the different selection test
 829 models.

Model		GR1 Sugar	GR2 Fructose	GR 3	GR4	GR5	GR6	GR7	GR8	GR9	GR10	GR11	GR12	
M0 vs M3	<i>InL</i>	-5231.825340	-1216.789767	-5789.152113	-9998.205966	-5479.083036	-6049.200897	-12089.821784	-9356.298793	-4445.435595	-9732.621983	-8102.433935	-4030.348531	
	<i>InL</i>	-5052.892647	-1211.381107	-5396.002334	-9300.606905	-5289.011093	-5710.513088	-11335.539722	-9049.122334	-4315.713162	-9222.209445	-7770.677152	-3953.943748	
	<i>2ΔL</i>	-357.87	10.817	786.3	1395.2	380.14	677.38	1508.6	614.35	259.44	1020.8	663.51	152.81	
	<i>K</i>	4	4	4	4	4	4	4	4	4	4	4	4	
	<i>alpha</i>	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
	<i>p</i>	2.2e-16	0.0287	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16
	χ^2	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877
M1a vs M2a	<i>InL</i>	-5077.671870	-1212.480402	-5490.213977	-9402.039709	-5353.305732	-5767.175345	-11420.967944	-9101.153882	-4366.772067	-9296.978281	-7794.902158	-3961.899495	
	<i>InL</i>	-5059.921483	-1211.381107	-5397.045087	-9304.423044	-5289.395383	-5713.611109	-11342.326713	-9053.610777	-4315.948000	-9222.323267	-7772.779594	-3953.982877	
	<i>2ΔL</i>	35.501	2.1986	186.34	195.23	127.82	107.13	157.28	95.086	101.65	149.31	44.245	15.833	
	<i>K</i>	2	2	2	2	2	2	2	2	2	2	2	2	
	<i>alpha</i>	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
	<i>p</i>	1.954811e-08	0.6992874	1	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.468e-10	0.0003646
	χ^2	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915
M7 vs M8	<i>InL</i>	-5077.945793	-1212.499290	-5480.009432	-9379.699699	-5365.565151	-5773.318298	-11401.784294	-9100.108902	-4367.195585	-9293.315424	-7804.778381	-3966.461479	
	<i>InL</i>	-5053.994057	-1211.382858	-5394.753409	-9290.460615	-5290.583339	-5716.276926	11327.505439	-9048.607034	-4316.436398	-9221.616983	-7771.030353	-3954.014079	
	<i>2ΔL</i>	47.903	2.2329	170.51	178.48	149.96	114.08	148.56	103	101.52	143.4	67.496	24.895	
	<i>K</i>	2	2	2	2	2	2	2	2	2	2	2	2	
	<i>alpha</i>	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
	<i>p</i>	3.961806e-11	0.327446	9.414119e-38	1.753712e-39	2.727802e-33	1.687512e-25	5.509385e-33	4.295591e-23	9.027452e-23	7.273823e-32	2.205046e-15	3.927922e-06	
	χ^2	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915

830 **Table S14 continuation.** Chemosensory genes (gustatory genes-GR and olfactory genes-OR) likelihood-rate test summary of the different
831 selection test models.

Model		OR1 Orco	OR2	OR3	OR4	OR5	OR6	OR7	OR8a	OR8b	OR9	OR10	OR11
M0 vs M3	<i>InL</i>	-1277.419980	-	-3324.042385	-	-3568.098925	-	-4479.693944	-3631.526679	-856.474056	-5319.972747	-	-
			4855.183909		3848.105608		3408.209158					3857.677939	2047.402217
	<i>InL</i>	-1251.614549	-	-3307.756934	-	-3495.220514	-	-4382.009169	-3522.071847	-786.305067	-5179.035727	-	-
			4779.219529		3801.408614		3275.335808					3799.432971	2031.469449
	2ΔL	51.611	151.93	32.571	93.394	145.76	265.75	195.37	218.91	140.34	281.87	116.49	31.866
	K	4	4	4	4	4	4	4	4	4	4	4	4
	<i>alpha</i>	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
p	1.663667e-10	2.2e-16	1.462175e-06	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.038e-06
χ²	9.4877	9.4877	9.4877	9.4877	-380.14	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877
M1a vs M2a	<i>InL</i>	-1259.699478	-	-3308.891909	-	-3543.684662	-3296.34482	-4396.215633	-3570.720139	-827.407331	-5267.011477	-	-
			4797.822208		3832.394693							3823.483550	2037.472383
	<i>InL</i>	-1251.776345	-	-3308.891909	-	-3497.405393	-	-4382.104041	-3522.928259	-786.311979	-5179.266618	-	-
			4781.144351		3801.445999		3276.383988					3799.778882	2031.484327
	2ΔL	15.846	33.356	0	61.897	92.559	39.922	28.223	95.584	82.191	175.49	47.409	11.976
	K	2	2	2	2	2	2	2	2	2	2	2	2
	<i>alpha</i>	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
p	0.0003623	5.713e-08	1	3.624e-14	7.964689e-21	2.143e-09	7.437e-07	1.754891e-21	1.420743e-18	7.814333e-39	5.072e-11	0.002509	
χ²	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915
M7 vs M8	<i>InL</i>	-1260.710668	-	-3308.277304	-	-3569.585274	-	-4396.427598	-3589.963693	-827.608477	-5267.344755	-	-
			4798.331056		3833.774262		3303.021239					3823.955115	2037.473281
	<i>InL</i>	-1251.866492	-	-3308.148161	-	-3497.460429	-	-4382.169578	-3523.414754	-786.377213	-5179.281417	-	-
			4781.370179		3801.432365		3279.169203					3799.798015	2031.476879
	2ΔL	17.688	33.922	0.25829	64.684	144.25	47.704	28.516	133.1	82.463	176.13	48.314	11.993
	K	2	2	2	2	2	2	2	2	2	2	2	2
	<i>alpha</i>	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
p	0.0001442	4.305e-08	0.8788	8.997e-15	2.2e-16	4.377e-11	6.424e-07	2.2e-16	2.2e-16	2.2e-16	3.226e-11	0.002488	
χ²	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915

832 **Table S14 continuation.** Chemosensory genes (gustatory genes-GR and olfactory genes-OR) likelihood-rate test summary of the different
 833 selection test models.

	Group 12	Group 13	Group 14	Group 15	Group 16	Group 17	Group 18	Group 19	Group 20	Group 21	Group 22	Group 23	
M0 vs M3	InL	-3588.139836	-3745.413691	-	-	-	-	-8552.735563	-	-6162.761013	-	-2678.888227	
	InL	-3437.771521	-3662.608215	2194.753612	5214.486356	2242.826138	4940.105104	-	8993.558702	2125.005660	3330.081615	-	
	2ΔL	300.74	165.61	2130.789706	5031.424098	2155.939738	4695.518331	-8097.321728	8659.143002	-5961.340929	2021.568078	3284.228800	-2573.756787
	K	4	4	4	4	4	4	4	4	4	4	4	4
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	p	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16
	χ²	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877
M1a vs M2a	InL	-3486.586634	-3681.069046	-	-	-	-	-8191.950929	-	-6020.326260	-	-2617.608997	
	InL	-3438.453620	-3662.868100	2158.354702	5063.053386	2186.389432	4711.510404	-	8678.778617	2036.359161	3304.977694	-	
	2ΔL	96.266	36.402	2131.706907	5031.428120	2156.215036	4701.791333	-8097.582281	8663.562640	-5963.486385	2021.667637	3284.332897	-2574.028379
	K	2	2	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	p	1.247664e-21	1.245746e-08	2.673e-12	1.842e-14	7.86e-14	6.013e-05	1.038047e-41	2.465e-07	2.06422e-25	4.164e-07	1.082e-09	1.183527e-19
	χ²	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915
M7 vs M8	InL	-3486.779859	-3684.323225	-	-	-	-	-8190.929871	-	-6020.805255	-	-2621.680496	
	InL	-3438.692997	-3662.634207	2164.038648	5071.873297	2186.708961	4703.931918	-	8685.315986	2034.553460	3305.056752	-	
	2ΔL	96.174	43.378	2131.892309	5031.964759	2156.051260	4693.116033	-8097.050353	8659.506895	-5963.772860	2021.029561	3284.481367	-2574.659541
	K	2	2	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	p	2.2e-16	3.807e-10	1.094e-14	2.2e-16	4.848e-14	2.008e-05	2.2e-16	6.184e-12	2.2e-16	1.339e-06	1.159e-09	2.2e-16
	χ²	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915

835 **Table S15.** Chemosensory genes (gustatory receptors GR and olfactory receptors OR) were
 836 suggested to evolve under positive and purifying selection based on the site model
 837 analysis.

Genes	Number of genes	Sites	Positively selected sites*		
			M0 vs M3	M1a vs M2a	M7 vs M8
GR01	44	354	M0	M1a	M7
<u>GR02</u>	<u>5</u>	<u>285</u>	<u>M0</u>	<u>M2a</u>	<u>M8</u>
GR03	56	234	M0	M2a	M7
GR04	57	408	M0	M1a	M7
GR05	25	519	M0	M1a	M7
GR06	32	456	M0	M1a	M7
GR07	67	441	M0	M1a	M7
GR08	24	822	M0	M1a	M7
GR09	22	441	M0	M1a	M7
GR10	32	639	M0	M1a	M7
GR11	19	858	M0	M1a	M7
GR12	10	789	M0	M1a	M7
Total	393	6246			

838

839

840 *Positive selection sites estimated under model M8 by Bayes empirical Bayes (BEB) approach with posterior probabilities (PPs) > 95%. The
 841 overlapping sites between M8 and M2a are underlined.

842

843

844

845

846

847

848

849

850

851

852

853 **Table S15 continuation.** Chemosensory genes (gustatory receptors GR and olfactory
 854 receptors OR) were suggested to evolve under positive and purifying selection based on
 855 the site model analysis.

856

Genes	Number of genes	Sites	M0 vs M3	M1a vs M2a	M7 vs M8	Positively Selected Sites
OR01 (Orco)	10	270	M0	M1a	M7	
OR02	11	1128	M0	M1a	M7	
OR03	5	<u>1104</u>	<u>M0</u>	<u>M2a</u>	<u>M8</u>	<u>345R</u>
OR04	9	795	M0	M1a	M7	
OR05	9	795	M0	M1a	M7	
OR06	37	312	M0	M1a	M7	
OR07	8	1002	M0	M1a	M7	
OR08a	15	444	M0	M1a	M7	
OR08b	31	60	M0	M1a	M7	
OR09	12	939	M0	M1a	M7	
OR10	10	1011	M0	M1a	M7	
OR11	8	579	M0	M1a	M7	
OR12	25	468	M0	M1a	M7	
OR13	11	885	M0	M1a	M7	
OR14	17	306	M0	M1a	M7	
OR15	14	678	M0	M1a	M7	
OR16	15	378	M0	M1a	M7	
OR17	25	387	M0	M1a	M7	
OR18	21	702	M0	M1a	M7	
OR19	21	1062	M0	M1a	M7	
OR20	14	1086	M0	M1a	M7	
OR21	15	195	M0	M1a	M7	
OR22	8	618	M0	M1a	M7	
OR23	22	300	M0	M1a	M7	
Total	373	15504				

857

858

*Positive selection sites estimated under model M8 by Bayes empirical Bayes (BEB) approach with posterior probabilities (PPs) > 95%. The overlapping sites between M8 and M2a are underlined.

859

860

861

862 **Table S16.** Selection test summary based on the branch model using single-copy orthologous groups were suggested under positive ($\omega > 0$),
 863 purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang, 2007).

864

865

866

867

868

869

870

871

872

	OG01	OG02	OG03	OG04	OG05	OG06	OG07	OG08	OG09	OG10
N	13	13	13	13	13	13	13	13	13	13
S	477	3060	1371	450	738	2679	1380	159	192	930
	2.24858	0.12296	0.50758	0.28491	0.49332	0.06533	0.91522	0.61627	2.26590	1.83966
lnL	-2160.159209	-12064.036465	-5078.987736	-3104.222491	-5169.454320	-10524.953899	-9353.560017	-763.114795	-556.016689	-6222.568903
lnL	-2151.066978	-12046.973192	-5060.117689	-3010.991489	-5144.933406	-10221.358852	-9340.036163	-754.339339	-552.390548	-6205.533917
2ΔL	18.184	34.127	37.74	186.46	170.02	607.19	27.048	17.551	7.2523	34.07
K	12	12	12	12	12	12	12	12	12	12
alpha	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
p	0.1102	0.0006442	0.0001693	2.2e-16	2.2e-16	2.2e-16	0.007606	0.13	0.8405	0.0006576
χ	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026

873

874

875

876

877

878 **Table S16 continuation.** Selection test summary based on the branch model using single-copy orthologous groups were suggested under
 879 positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang, 2007).

880

	OG11	OG12	OG13	OG14	OG15	OG16	OG17	OG18	OG19	OG20
<i>N</i>	13	13	13	13	13	13	13	13	13	13
<i>S</i>	1617	1512	474	2178	813	945	1296	1005	678	939
	1.22681	0.09122	2.69628	0.31425	0.03437	0.09813	0.10343	0.06158	0.06811	0.30141
<i>InL</i>	-13934.579561	-6346.701804	-2279.718515	-12827.691404	-2882.013677	-3921.725208	-5865.130797	-3778.694526	-2451.561446	-5439.865077
<i>InL</i>	-13926.456658	-6329.805317	-2268.388919	-12797.149749	-2871.227283	-3910.795705	-5588.520143	-3746.812685	-2438.768897	-5407.142548
<i>2ΔL</i>	16.246	33.793	22.659	61.083	21.573	21.859	553.22	63.764	25.585	65.445
<i>K</i>	12	12	12	12	12	12	12	12	12	12
<i>alpha</i>	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
<i>p</i>	0.1802	0.0007271	0.03076	1.432e-08	0.0426	0.03913	2.2e-16	4.611e-09	0.01228	2.256e-09
<i>χ</i>	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026

881

882

883

884

885

886

887

888 **Table S16 continuation.** Selection test summary based on the branch model using single-copy orthologous groups were suggested under
 889 positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang, 2007).

890

	OG21	OG22	OG23	OG24	OG25	OG26	OG27	OG28	OG29	OG30
N	13	13	13	13	13	13	13	13	13	13
S	1317	537	435	600	909	336	606	1575	477	702
	3.27177	0.06587	0.25804	0.02001	0.29891	0.05698	0.63361	0.38606	2.24858	0.23299
lnL	-	-1972.640187	-2845.469017	-1942.623794	-6987.667834	-1181.017753	-4152.181626	-11145.162243	-3419.975689	-4661.712646
	7167.360574									
lnL	-	-1959.182468	-2799.426152	-1927.852835	-6868.104794	-1171.684502	-4118.493890	-10923.892619	-3299.384156	-4457.970644
	7127.128098									
2ΔL	80.465	26.915	92.086	29.542	239.13	18.667	67.375	442.54	241.18	407.48
K	12	12	12	12	12	12	12	12	12	12
alph	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
a										
p	3.365e-12	0.007947	1.946e-14	0.003271	2.2e-16	0.0969	9.888e-10	2.2e-16	2.2e-16	2.2e-16
χ	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026

891

892

893

894

895

896

897 **Table S16 continuation.** Selection test summary based on the branch model using single-copy orthologous groups were suggested under
 898 positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang, 2007).

899

	OG31	OG32	OG33	OG34	OG35	OG36	OG37	OG38	OG39	OG40
N	13	13	13	13	13	13	13	13	13	13
S	1527	435	1854	1800	1029	879	1707	480	840	1425
	0.46942	1.12776	0.78526	0.22210	2.81737	0.04742	0.27643	0.93398	0.24967	0.13977
InL	- 8759.94630 3	-2848.854168	-14544.838549	-8886.695081	-3913.272980	-2874.368149	-13163.628869	-3897.522790	-5862.293072	-6962.314369
InL	- 8751.41631 7	-2840.262743	-14508.528505	-8867.252088	-3902.348755	-2858.457621	-12674.445587	-3885.905342	-5699.290353	-6772.689588
2ΔL	17.06	17.183	72.62	38.886	21.848	31.821	978.37	23.235	326.01	379.25
K	12	12	12	12	12	12	12	12	12	12
alph a	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
p	0.1474	0.1428	1.033e-10	0.0001099	0.03925	0.001475	2.2e-16	0.0258	2.2e-16	2.2e-16
χ	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026

900

901

902

903

904

905

906 **Table S16 continuation.** Selection test summary based on the branch model using single-copy orthologous groups were suggested under
 907 positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang, 2007).

908

	OG41	OG42	OG43	OG44	OG45	OG46	OG47	OG48	OG49	OG50
N	13	13	13	13	13	13	13	13	13	13
S	1425	609	795	912	1434	948	696	297	1806	495
	3.00446	1.16905	4.06583	5.07984	0.21805	0.02876	0.23508	1.81814	0.03415	0.11660
InL	- 1708.725361	-2766.130509	-3105.508058	-3766.866329	-9067.084798	-3342.235869	-3052.410174	-1334.131281	-6072.369179	-2532.568823
InL	- 1695.593096	-2755.461167	-3092.621569	-3756.804592	-8838.336867	-3317.490890	-3035.356911	-1324.903251	-6057.761169	-2478.000576
2ΔL	26.265	21.339	25.773	20.123	457.5	49.49	34.107	18.456	29.216	109.14
K	12	12	12	12	12	12	12	12	12	12
alph a	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
p	0.009846	0.04564	0.01156	0.06479	2.2e-16	1.717e-06	0.0006489	0.1025	0.003659	2.2e-16
χ	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026	21.026

909

910

911

912

913

914

915 **Table S17.** Single-copy orthologous groups (OG) were suggested under positive ($\omega > 0$),
 916 purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4
 917 (Yang, 2007).

918

Genes	Number of genes	Sites	Free Omega	$\omega > 0$	$\omega < 0$	$\omega = 0$	Saturated
OG01	13	477	No	4	1	0	8
OG02	13	3060	Yes	0	13	0	0
OG03	13	1371	Yes	1	11	0	1
OG04	13	450	Yes	1	11	0	1
OG05	13	738	Yes	1	11	0	1
OG06	13	2679	Yes	0	13	0	0
OG07	13	1380	Yes	2	9	0	2
OG08	13	159	No	1	8	0	4
OG09	13	192	No	4	3	0	6
OG10	13	930	Yes	4	2	0	7
OG11	13	1617	No	10	2	0	1
OG12	13	1512	Yes	0	13	0	0
OG13	13	474	Yes	4	1	0	8
OG14	13	2178	Yes	1	12	0	0
OG15	13	813	Yes	0	13	0	0
OG16	13	945	Yes	0	13	0	0
OG17	13	1296	Yes	<u>0</u>	<u>13</u>	<u>0</u>	0
OG18	13	1005	Yes	1	12	0	0
OG19	13	678	Yes	0	13	0	0
OG20	13	939	Yes	0	13	0	0
OG21	13	1317	Yes	<u>0</u>	<u>3</u>	<u>0</u>	10
OG22	13	537	Yes	<u>1</u>	<u>12</u>	<u>0</u>	0
OG23	13	435	Yes	<u>0</u>	<u>11</u>	<u>0</u>	2
OG24	13	600	Yes	<u>0</u>	<u>12</u>	<u>0</u>	1
OG25	13	909	Yes	<u>1</u>	<u>12</u>	<u>0</u>	0
OG26	13	336	Yes	<u>0</u>	<u>13</u>	<u>0</u>	0
OG27	13	606	Yes	2	11	0	0
OG28	13	1575	Ha	0	11	0	2
OG29	13	477	Yes	2	9	0	2
OG30	13	702	Yes	<u>1</u>	<u>12</u>	<u>0</u>	0

919

920

921

922 **Table S17 continuation.** Single-copy orthologous groups (OG) were suggested under positive
 923 ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML
 924 v4 (Yang, 2007).

925

Genes	Number of genes	Sites	Free Omega	$\omega > 0$	$\omega < 0$	$\omega = 0$	Saturated
OG31	13	1527	No	1	12	0	0
OG32	13	435	No	6	2	0	5
OG33	13	1854	Yes	2	11	0	0
OG34	13	1800	Yes	0	13	0	0
OG35	13	1029	Yes	2	1	0	10
OG36	13	879	Yes	0	13	0	0
OG37	13	1707	Yes	1	12	0	0
OG38	13	480	Yes	3	9	0	1
OG39	13	840	Yes	0	13	0	0
OG40	13	1425	Yes	1	12	0	0
OG41	13	1425	Yes	3	1	0	9
OG42	13	609	Yes	5	4	0	4
OG43	13	795	Yes	1	0	0	12
OG44	13	912	No	1	0	0	12
OG45	13	1434	Yes	1	12	0	0
OG46	13	948	Yes	0	13	0	0
OG47	13	696	Yes	1	10	0	2
OG48	13	297	Yes	1	4	0	8
OG49	13	1806	Yes	0	13	0	0
OG50	13	495	Yes	0	13	0	0
Total	650	51780		70	461	0	119

926

927 **Table S18.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous groups were
 928 suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang, 2007).

929

Model	Statistics	OG01	OG02	OG03	OG04	OG05	OG06	OG07	OG08	OG09	OG10
M0	<i>N</i>	13	13	13	13	13	13	13	13	13	13
	<i>S</i>	477	3060	1371	450	738	2679	1380	159	192	930
	ω	2.03903	0.12625	0.49153	0.31310	0.58051	0.06688	0.96506	0.63003	1.98810	2.06902
	<i>lnL</i>	-2162.948435	-12069.630996	-5080.833417	-3121.452981	-5223.453997	-10548.369935	-9359.540371	-763.252274	-557.219077	-6236.572796
M1a	<i>p0</i>	0.25982	0.93998	0.66882	0.84218	0.77591	0.96046	0.69863	0.69878	0.00001	0.68480
	<i>p1</i>	0.74018	0.06002	0.33118	0.15782	0.22409	0.03954	0.30137	0.30122	0.99999	0.31520
	<i>lnL</i>	-2169.235915	-11808.787819	-4986.978991	-3034.793061	-4977.844425	-10364.330361	-9050.292984	-748.377623	-558.740901	-6083.510086
M2a	<i>p0</i>	0.53529	0.93998	0.70344	0.84029	0.76595	0.96046	0.65404	0.77167	0.48256	0.62337
	<i>p1</i>	0.33891	0.06002	0.25842	0.15850	0.19908	0.00841	0.27025	0.18466	0.00000	0.21548
	<i>p2</i>	0.12580	0.00000	0.03814	0.00122	0.03497	0.03113	0.07571	0.04367	0.51744	0.16115
	$\omega 2$	7.63607	27.59557	3.00940	4.36479	2.65723	1.00000	3.64310	3.88546	4.01022	5.72098
	<i>lnL</i>	-2120.742582	-11808.787825	-4977.452908	-3033.125159	-4963.865531	-10364.330360	-8947.296962	-743.102722	-553.225759	-5937.927170
M3	<i>p0</i>	0.83666	0.84481	0.77514	0.70166	0.76994	0.87104	0.70207	0.87603	0.48256	0.70541
	<i>p1</i>	0.16304	0.14619	0.21972	0.29002	0.20264	0.01563	0.24234	0.12193	0.11207	0.19961
	<i>p2</i>	0.00030	0.00900	0.00515	0.00831	0.02742	0.11333	0.05559	0.00204	0.40537	0.09499
	$\omega 0$	0.96768	0.01450	0.06293	0.00000	0.00000	0.00000	0.04499	0.19065	0.00000	0.01740
	$\omega 1$	6.52139	0.41428	1.45284	0.54568	1.06355	0.00000	1.28657	2.27276	4.01022	2.43871
	$\omega 2$	82.49708	1.96107	5.42739	2.60606	2.84132	0.42166	4.04399	11.24185	4.01022	6.94049
	<i>lnL</i>	-2112.487363	-11800.164060	-4977.163013	-3028.144815	-4963.837256	-10328.578162	-8946.887586	-742.909191	-553.225759	-5937.036209

930 **Table S18 continuation.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous
 931 groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang,
 932 2007).

933

Model	Statistics	OG11	OG12	OG13	OG14	OG15	OG16	OG17	OG18	OG19	OG20
M0	<i>N</i>	13	13	13	13	13	13	13	13	13	13
	<i>S</i>	1617	1512	474	2178	813	945	1296	1005	678	939
	ω	1.27020	0.09304	2.51172	0.31578	0.03617	0.09932	0.10720	0.06236	0.10720	0.06236
	<i>lnL</i>	-14004.587976	-6347.668937	-2282.294164	-12827.962628	-2885.791887	-3922.364464	-5887.196356	-3779.892485	-5887.196356	-3779.892485
M1a	<i>p0</i>	0.63453	0.95844	0.25594	0.83268	0.98920	0.95672	0.92350	0.96656	0.92350	0.96656
	<i>p1</i>	0.36547	0.04156	0.74406	0.16732	0.01080	0.04328	0.07650	0.03344	0.07650	0.03344
	<i>lnL</i>	-13663.361889	-6211.419555	-2297.410342	-12519.847267	-2836.443831	-3834.113964	-5765.926799	-3693.393850	-5765.926799	-3693.393850
M2a	<i>p0</i>	0.59730	0.95844	0.49500	0.83268	0.99005	0.95672	0.92350	0.96656	0.92350	0.96656
	<i>p1</i>	0.29631	0.04156	0.29370	0.09452	0.00926	0.04328	0.04754	0.01417	0.04754	0.01417
	<i>p2</i>	0.10640	0.00000	0.21130	0.07280	0.00069	0.00000	0.02896	0.01926	0.02896	0.01926
	$\omega 2$	3.78312	29.83620	30.60936	1.00000	2.17707	35.25915	1.00000	1.00000	1.00000	1.00000
	<i>lnL</i>	-13467.529096	-6211.419555	-2245.293250	-12519.847267	-2836.296475	-3834.113964	-5765.926798	-3693.393850	-5765.926798	-3693.393850
M3	<i>p0</i>	0.65100	0.21929	0.68509	0.57719	0.67016	0.75483	0.87749	0.90476	0.87749	0.90476
	<i>p1</i>	0.30146	0.72816	0.31003	0.35780	0.32330	0.22457	0.07794	0.09465	0.07794	0.09465
	<i>p2</i>	0.04754	0.05255	0.00488	0.06500	0.00653	0.02060	0.04457	0.00059	0.04457	0.00059
	$\omega 0$	0.00855	0.02486	0.92539	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	$\omega 1$	1.57218	0.02486	5.61221	0.33608	0.05044	0.19885	0.60403	0.39796	0.60403	0.39796
	$\omega 2$	5.02400	0.80591	30.60936	1.43687	1.32205	1.11094	0.60404	4.69666	0.60404	4.69666
	<i>lnL</i>	-13464.492160	-6209.966418	-2241.487657	-12510.085125	-2836.115816	-3825.078218	-5754.333055	-3682.333934	-5754.333055	-3682.333934

934 **Table S18 continuation.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous
 935 groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang,
 936 2007).

Model	Statistics	OG21	OG22	OG23	OG24	OG25	OG26	OG27	OG28	OG29	OG30
M0	<i>N</i>	13	13	13	13	13	13	13	13	13	13
	<i>S</i>	1317	537	435	600	909	336	606	1575	477	702
	ω	3.15645	0.06869	0.32988	0.02128	0.37197	0.05701	0.75611	0.48197	0.17739	0.24196
	<i>lnL</i>	-7173.277608	-1973.476058	-2893.070677	-1945.235038	-7079.463353	-1181.331501	-4165.726469	-11255.347165	-3447.167177	-4718.510529
M1a	<i>p0</i>	0.34621	0.97365	0.87254	0.98696	0.84245	0.98405	0.76337	0.80323	0.89282	0.94565
	<i>p1</i>	0.65379	0.02635	0.12746	0.01304	0.15755	0.01595	0.23663	0.19677	0.10718	0.05435
	<i>lnL</i>	-7228.052639	-1943.028431	-2749.813253	-1919.354366	-6804.813715	-1169.588702	-4009.024521	-10796.382824	-3363.304373	-4647.758019
M2a	<i>p0</i>	0.25129	0.97365	0.87022	0.98696	0.83854	0.98405	0.74170	0.79261	0.89281	0.94565
	<i>p1</i>	0.53641	0.02635	0.12640	0.00219	0.15942	0.00105	0.21948	0.19533	0.10719	0.01669
	<i>p2</i>	0.21230	0.00000	0.00338	0.01084	0.00204	0.01490	0.03882	0.01206	0.00000	0.03766
	$\omega 2$	8.43115	43.72036	3.61611	1.00000	4.04644	1.00000	3.84791	3.81021	13.80350	1.00000
	<i>lnL</i>	-7015.381481	-1943.028431	-2747.641931	-1919.354367	-6801.142675	-1169.588702	-3981.442258	-10767.549294	-3363.304385	-4647.758019
M3	<i>p0</i>	0.52851	0.52210	0.8101	0.03091	0.65664	0.11535	0.79904	0.76887	0.74728	0.56162
	<i>p1</i>	0.37439	0.45088	0.18157	0.95034	0.32681	0.86845	0.18712	0.21466	0.25199	0.32317
	<i>p2</i>	0.09710	0.02702	0.00831	0.01876	0.01655	0.01621	0.01384	0.01647	0.00073	0.11521
	$\omega 0$	0.48277	0.00000	0.02082	0.00000	0.00000	0.02986	0.04974	0.01540	0.00000	0.00000
	$\omega 1$	3.44893	0.05436	0.66638	0.00000	0.45069	0.02986	1.52866	0.88168	0.44942	0.40801
	$\omega 2$	11.61563	0.90949	2.83159	0.72032	2.10823	0.98764	5.53708	3.46481	3.96368	0.40801
	<i>lnL</i>	-7012.525052	-1942.808084	-2746.114847	-1918.986824	-6782.359176	-1169.588354	-3980.362094	-10767.412053	-3347.711898	-4625.850279

937

938 **Table S18 continuation.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous
 939 groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang,
 940 2007).

Model	Statistics	OG31	OG32	OG33	OG34	OG35	OG36	OG37	OG38	OG39	OG40
M0	<i>N</i>	13	13	13	13	13	13	13	13	13	13
	<i>S</i>	1527	435	1854	1800	1029	879	1707	480	840	1425
	ω	0.45877	1.34443	0.92485	0.23130	2.36009	0.04643	0.32492	1.13044	0.28521	0.15448
	<i>lnL</i>	-8768.449410	-2868.612793	-14621.285148	-8894.045372	-3929.296093	-2874.910280	-13312.710408	-3928.586702	-5930.825560	-7009.639226
M1a	<i>p0</i>	0.76233	0.73731	0.68654	0.89133	0.17106	0.98330	0.92134	0.72567	0.91023	0.92920
	<i>p1</i>	0.23767	0.26269	0.31346	0.10867	0.82894	0.01670	0.07866	0.27433	0.08977	0.07080
	<i>lnL</i>	-8564.647612	-2752.368582	-14140.983283	-8644.277513	-3942.697472	-2853.656918	-13093.487688	-3756.388973	-5805.481261	-6890.667418
M2a	<i>p0</i>	0.75935	0.65871	0.63893	0.88985	0.21863	0.98330	0.92134	0.66751	0.91022	0.92920
	<i>p1</i>	0.23161	0.28368	0.30690	0.10982	0.61608	0.00645	0.05631	0.26635	0.04861	0.07080
	<i>p2</i>	0.00904	0.05761	0.05417	0.00033	0.16529	0.01025	0.02235	0.06614	0.04116	0.00000
	$\omega 2$	3.98219	5.98054	3.35478	9.17946	8.67694	1.00000	1.00000	3.93325	1.00000	16.97923
	<i>lnL</i>	-8545.108356	-2690.644641	-14061.772045	-8635.257293	-3856.077852	-2853.656918	-13093.487688	-3712.184796	-5805.481261	-6890.667420
M3	<i>p0</i>	0.77399	0.74277	0.59986	0.82584	0.58799	0.79781	0.49340	0.75091	0.53446	0.70661
	<i>p1</i>	0.21901	0.24127	0.32875	0.16955	0.37949	0.19925	0.50364	0.23725	0.46308	0.28955
	<i>p2</i>	0.00700	0.01596	0.07139	0.00462	0.03252	0.00294	0.00296	0.01184	0.00246	0.00384
	$\omega 0$	0.08158	0.00072	0.00097	0.03691	0.83729	0.00000	0.00000	0.03166	0.00000	0.00000
	$\omega 1$	1.06919	2.21512	0.83809	0.63693	4.26197	0.16510	0.41658	1.88447	0.37786	0.34956
	$\omega 2$	4.33771	10.76177	3.05736	3.06968	20.76164	2.04210	2.90455	7.17490	4.53184	1.99404
	<i>lnL</i>	-8544.994586	-2687.374379	-14061.496172	-8630.576510	-3850.677314	-2853.074629	-	-3710.943836	-5773.580811	-6872.342203
							13065.035639				

941

942 **Table S18 continuation.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous
 943 groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang,
 944 2007).

945

Model	Statistics	OG41	OG42	OG43	OG44	OG45	OG46	OG47	OG48	OG49	OG50
M0	N	13	13	13	13	13	13	13	13	13	13
	S	1425	609	795	912	1434	948	696	297	1806	495
	ω	3.01909	1.15538	3.53209	4.26656	0.25442	0.03075	0.24336	1.64748	0.03449	0.14152
	lnL	-1708.733339	-2773.327806	-3118.301126	-3785.486916	-9144.155714	-3347.179895	-3053.244689	-1345.203631	-6073.114803	-2544.769522
M1a	p0	0.33890	0.35150	0.06555	0.00001	0.89824	0.98803	0.88764	0.29450	0.98742	0.92757
	p1	0.66110	0.64850	0.93445	0.99999	0.10176	0.01197	0.11236	0.70550	0.01258	0.07243
	lnL	-1719.991061	-2753.946889	-3152.415326	-3843.277682	-8919.315424	-3303.994599	-3004.787145	-1335.653420	-6008.522814	-2505.373631
M2a	p0	0.37669	0.39400	0.71781	0.03567	0.89824	0.98803	0.89076	0.75688	0.98742	0.92757
	p1	0.44150	0.33095	0.02892	0.67106	0.04521	0.01197	0.10743	0.00000	0.00385	0.00796
	p2	0.18181	0.27505	0.25328	0.29327	0.05655	0.00000	0.00181	0.24312	0.00873	0.06447
	$\omega 2$	7.16173	2.37170	8.59261	10.38207	1.00000	1.00000	3.06773	4.20740	1.00000	1.00000
	lnL	-1676.231683	-2741.944017	-3069.716063	-3713.752060	-8919.315424	-3303.994599	-3004.514350	-1315.858938	-6008.522814	-2505.373631
M3	p0	0.32369	0.42989	0.72723	0.16197	0.63657	0.98235	0.83758	0.70683	0.00002	0.36931
	p1	0.61593	0.55964	0.21808	0.70016	0.35649	0.00000	0.14373	0.05005	0.98385	0.41799
	p2	0.06038	0.01046	0.05469	0.13786	0.00694	0.01765	0.01870	0.24312	0.01613	0.21270
	$\omega 0$	0.00000	0.00000	1.80838	0.00000	0.00000	0.00735	0.06465	0.59565	0.01229	0.00615
	$\omega 1$	2.80261	1.64296	6.37883	3.08324	0.37640	0.04821	0.65885	0.59565	0.01324	0.00621
	$\omega 2$	12.47007	6.90660	19.93826	17.55211	2.23659	0.71720	1.82138	4.20740	0.82041	0.39239
	lnL	-1673.816166	-2741.298992	-3064.991199	-3709.550828	-8887.890899	-3303.396341	-3004.355660	-1315.858938	-6008.146911	-2494.922373

946 **Table S18 continuation.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous
 947 groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang,
 948 2007).

949

Model	Statistics	OG01	OG02	OG03	OG04	OG05	OG06	OG07	OG08	OG09	OG10
M7	<i>p</i>	3.99238	0.07707	0.00581	0.09053	0.00500	0.02946	0.01058	0.00563	7.28889	0.00500
	<i>q</i>	0.00500	0.72204	0.01124	0.42311	0.01944	0.53253	0.02130	0.00707	0.00500	0.01165
	<i>lnL</i>	-2174.312191	-11814.932584	-4987.502003	-3032.430812	-4978.958372	-10329.253345	-9050.732544	-748.936935	-558.740903	-6083.778050
M8	<i>p0</i>	0.87420	0.99124	0.95553	0.99660	0.96159	0.99929	0.92585	0.93849	0.48256	0.83322
	<i>p1</i>	0.12580	0.00876	0.04447	0.00340	0.03841	0.00071	0.07415	0.06151	0.51744	0.16678
	<i>p</i>	2.75653	0.12909	0.05686	0.12628	0.00502	0.02987	0.01847	0.58485	0.00500	0.01733
	<i>q</i>	0.00500	1.50574	0.14691	0.60838	0.02101	0.55288	0.03754	1.55468	2.06968	0.05382
	ω	7.63607	1.96894	2.89305	3.26448	2.57060	1.61632	3.69431	3.43188	4.01022	5.63760
	<i>lnL</i>	-2120.742584	-11800.268657	-4977.528737	-3029.343683	-4963.886655	-10328.535146	-8947.455539	-743.192940	-553.225759	-5938.151822

950

951

952

953

954

955

956 **Table S18 continuation.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous
 957 groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang,
 958 2007).

959

Model	Statistics	OG11	OG12	OG13	OG14	OG15	OG16	OG17	OG18	OG19	OG20
M7	<i>p</i>	0.00500	0.07424	64.77837	0.12198	0.04111	0.07992	0.01493	0.01399	0.01493	0.01399
	<i>q</i>	0.00749	0.91728	0.00500	0.46995	1.04170	0.97517	0.23436	0.27037	0.23436	0.27037
	<i>lnL</i>	- 13665.327862	-6218.707134	-2303.037618	-12521.830412	-2846.434892	-3829.332524	-5755.809986	-3688.029348	- 5755.809986	-3688.029348
M8	<i>p0</i>	0.88231	0.96369	0.78870	0.97109	0.99347	0.98088	0.99999	0.99938	0.99999	0.99938
	<i>p1</i>	0.11769	0.03631	0.21130	0.02891	0.00653	0.01912	0.00001	0.00062	0.00001	0.00062
	<i>p</i>	0.00500	0.46232	98.99668	0.20687	0.31352	0.13805	0.01230	0.00550	0.01230	0.00550
	<i>q</i>	0.01164	13.24185	0.00500	1.01252	17.22106	2.50834	0.19903	0.10876	0.19903	0.10876
	ω	3.61834	1.00000	6.86935	1.76407	1.32106	1.11066	1.69552	4.60556	1.69552	4.60556
	<i>lnL</i>	- 13467.984638	-6210.890798	-2245.293250	-12510.567346	-2836.132958	-3825.849050	-5755.813062	-3682.383972	- 5755.813062	-3682.383972

960

961

962

963

964

965

966 **Table S18 continuation.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous
 967 groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang,
 968 2007).

969

Model	Statistics	OG21	OG22	OG23	OG24	OG25	OG26	OG27	OG28	OG29	OG30
M7	<i>p</i>	2.14132	0.06406	0.06593	0.01157	0.11791	0.09797	0.00500	0.01324	0.12438	0.39905
	<i>q</i>	0.00500	1.03157	0.37167	0.29128	0.55212	1.71366	0.01043	0.04401	0.91871	1.75626
	<i>lnL</i>	-7260.652141	-1945.359118	-2750.956795	-1925.351999	-6793.559395	-1171.791179	-4011.459227	-10797.123547	-3351.699957	-4631.216856
M8	<i>p0</i>	0.78768	0.97751	0.99254	0.98709	0.99358	0.98435	0.95524	0.98419	0.99947	0.99980
	<i>p1</i>	0.21232	0.02249	0.00746	0.01291	0.00642	0.01565	0.04476	0.01581	0.00053	0.00020
	<i>p</i>	3.80620	0.37626	0.11086	0.01337	0.17085	3.12902	0.04091	0.05149	0.13684	0.42276
	<i>q</i>	0.00500	12.46098	0.66878	1.53760	0.83982	99.00000	0.13223	0.19849	1.02612	1.87258
	ω	8.43172	1.00000	2.88271	1.00000	2.75890	1.00000	3.66200	3.49762	4.13548	5.57258
	<i>lnL</i>	-7015.381484	-1942.826664	-2746.445973	-1919.346836	-6785.007090	-1169.662614	-3981.889340	-10767.435440	-3350.794888	-4630.941590

970

971

972

973

974

975

976 **Table S18 continuation.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous
 977 groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang,
 978 2007).

979

Model	Statistics	OG31	OG32	OG33	OG34	OG35	OG36	OG37	OG38	OG39	OG40
M7	<i>p</i>	0.09682	0.00500	0.01361	0.12853	2.93311	0.05823	0.42458	0.00500	0.27664	0.15619
	<i>q</i>	0.23303	0.01169	0.02728	0.71945	0.00500	1.15745	1.52297	0.01164	1.18964	1.22612
	<i>lnL</i>	-8574.059132	-2753.230060	-14141.789697	-8646.652568	-3947.301727	-2855.538348	-13079.471158	-3756.977191	-5794.915321	-6875.501372
M8	<i>p0</i>	0.97856	0.94224	0.94255	0.99376	0.83471	0.99740	0.99922	0.93748	0.99805	0.99865
	<i>p1</i>	0.02144	0.05776	0.05745	0.00624	0.16529	0.00260	0.00078	0.06252	0.00195	0.00135
	<i>p</i>	0.19284	0.00500	0.02227	0.18461	5.34057	0.11980	0.51350	0.00500	0.43635	0.17342
	<i>q</i>	0.53024	0.01181	0.04396	1.14440	0.00500	2.99264	1.86388	0.01182	1.94329	1.40620
	ω	2.99660	5.96985	3.27229	2.75081	8.67694	2.14111	4.67567	4.05177	4.92497	2.42542
	<i>lnL</i>	-8547.602397	-2690.644795	-14061.493950	-8630.991740	-3856.077856	-2853.096431	-13070.395831	-3712.245453	-5779.817707	-6874.751966

980

981

982

983

984

985

986 **Table S18 continuation.** Single-copy orthologous groups (OG) selection test summary based on the branch model using single-copy orthologous
 987 groups were suggested under positive ($\omega > 0$), purifying ($\omega < 0$), and neutral ($\omega = 0$) selection based on the branch model analysis PAML v4 (Yang,
 988 2007).

989

Model	Statistics	OG41	OG42	OG43	OG44	OG45	OG46	OG47	OG48	OG49	OG50
M7	<i>p</i>	2.75291	2.75463	13.81559	5.88250	0.17780	0.01270	0.16541	18.57656	0.02479	0.13611
	<i>q</i>	0.00500	0.00500	0.00500	0.00500	0.98920	0.30150	0.73963	0.00500	0.60206	1.34146
	<i>lnL</i>	-1727.628142	-2767.010532	-3152.738153	-3843.277680	-8900.392501	-3308.237776	-3007.209124	-1340.019158	-6016.002738	-2495.413217
M8	<i>p0</i>	0.81819	0.68841	0.74672	0.70673	0.99690	0.99170	0.95382	0.75717	0.98870	0.99999
	<i>p1</i>	0.18181	0.31159	0.25328	0.29327	0.00310	0.00830	0.04618	0.24283	0.01130	0.00001
	<i>p</i>	1.68251	0.01415	2.73622	1.08685	0.23483	0.01760	0.48285	99.00000	0.35122	0.13612
	<i>q</i>	0.00500	0.01909	0.00500	0.00500	1.37695	0.63598	3.34051	66.97474	20.92990	1.34165
	ω	7.16174	2.26902	8.59261	10.38204	2.77638	1.00000	1.43038	4.20947	1.00000	1.00000
	<i>lnL</i>	-1676.231688	-2741.965899	-3069.716070	-3713.752084	-8891.840868	-3303.653754	-3004.453461	-1315.860983	-6008.361196	-2495.413624

990

991

992

993

994

995

996 **Table S19.** Single-copy orthologous groups (OG) likelihood-rate test summary of the different selection test models.

Model		OG01	OG02	OG03	OG04	OG05	OG06	OG07	OG08	OG09	OG10	
M0 vs M3	lnL	-2162.948435	-12069.630996	-5080.833417	-3121.452981	-5223.453997	-10548.369935	-9359.540371	-763.252274	-557.219077	-6236.572796	
	lnL	-2112.487363	-11800.164060	-4977.163013	-3028.144815	-4963.837256	-10328.578162	-8946.887586	-742.909191	-553.225759	-5937.036209	
	2ΔL	100.92	538.93	207.34	186.62	519.23	439.58	825.31	40.686	7.9866	599.07	
	K	4	4	4	4	4	4	4	4	4	4	
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
	p	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	3.122e-08	0.09207	2.2e-16
	χ	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	
M1a vs M2a	lnL	-2169.235915	-11808.787819	-4986.978991	-3034.793061	-4977.844425	-10364.330361	-9050.292984	-748.377623	-558.740901	-6083.510086	
	lnL	-2120.742582	-11808.787825	-4977.452908	-3033.125159	-4963.865531	-10364.330360	-8947.296962	-743.102722	-553.225759	-5937.927170	
	2ΔL	96.987	-1.2e-05	19.052	3.3358	27.958	2,00E-06	205.99	10.55	11.03	291.17	
	K	2	2	2	2	2	2	2	2	2	2	
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
	p	2.2e-16	1	7.292e-05	0.1886	8.493e-07	1	2.2e-16	0.005118	0.004025	2.2e-16	
	χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	
M7 vs M8	lnL	-2174.312191	-11814.932584	-4987.502003	-3032.430812	-4978.958372	-10329.253345	-9050.732544	-748.936935	-558.740903	-6083.778050	
	lnL	-2120.742584	-11800.268657	-4977.528737	-3029.343683	-4963.886655	-10328.535146	-8947.455539	-743.192940	-553.225759	-5938.151822	
	2ΔL	107.14	29.328	19.947	6.1743	30.143	1.4364	206.55	11.488	11.03	291.25	
	K	2	2	2	2	2	2	2	2	2	2	
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
	p	2.2e-16	4.281e-07	4.663e-05	0.04563	2.847e-07	0.4876	2.2e-16	0.003202	0.004025	2.2e-16	
	χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	

997

998 **Table S19 continuation.** Single-copy orthologous groups (OG) likelihood-rate test summary of the different selection test models.

Model	OG11	OG12	OG13	OG14	OG15	OG16	OG17	OG18	OG19	OG20	
M0 vs M3	InL	-	-	-	-	-	-	-	-	-	
		14004.587976	-6347.668937	-2282.294164	12827.962628	-2885.791887	-3922.364464	-5887.196356	3779.892485	5887.196356	-3779.892485
	InL	-	-	-	-	-	-	-	-	-	
		13464.492160	-6209.966418	-2241.487657	12510.085125	-2836.115816	-3825.078218	-5754.333055	3682.333934	5754.333055	-3682.333934
	2ΔL	1080.2	275.41	81.613	635.76	99.352	194.5	265.73	195.12	265.73	195.12
	K	4	4	4	4	4	4	4	4	4	4
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
p	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	
χ	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	
M1a vs M2a	InL	-	-	-	-	-	-	-	-	-	
		13663.361889	-6211.419555	-2297.410342	12519.847267	-2836.443831	-3834.113964	-5765.926799	3693.393850	5765.926799	-3693.393850
	InL	-	-	-	-	-	-	-	-	-	
		13467.529096	-6211.419555	-2245.293250	12519.847267	-2836.296475	-3834.113964	-5765.926798	3693.393850	5765.926798	-3693.393850
	2ΔL	391.67	0	104.23	0	0.29471	0	2,00E-06	0	2,00E-06	0
	K	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
P	2.2e-16	1	2.2e-16	1	0.863	1	1	1	1	1	
χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	
M7 vs M8	InL	-	-	-	-	-	-	-	-	-	
		13665.327862	-6218.707134	-2303.037618	12521.830412	-2846.434892	-3829.332524	-5755.809986	3688.029348	2406.888553	-3688.029348
	InL	-	-	-	-	-	-	-	-	-	
		13467.984638	-6210.890798	-2245.293250	12510.567346	-2836.132958	-3825.849050	-5755.813062	3682.383972	2396.599605	-3682.383972
	2ΔL	394.69	15.633	115.49	22.526	20.604	6.9669	-0.006152	11.291	20.578	11.291
	K	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
P	2.2e-16	0.0004031	2.2e-16	1.284e-05	3.357e-05	0.0307	1	0.003534	3.401e-05	0.003534	
χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	

1000 **Table S19 continuation.** Single-copy orthologous groups (OG) likelihood-rate test summary of the different selection test models

Model	OG21	OG22	OG23	OG24	OG25	OG26	OG27	OG28	OG29	OG30	
M0 vs M3	lnL	-7173.277608	-1973.476058	-2893.070677	-1945.235038	-7079.463353	-1181.331501	-4165.726469	-	-	
	lnL	-7012.525052	-1942.808084	-2746.114847	-1918.986824	-6782.359176	-1169.588354	-3980.362094	10767.412053	3347.711898	
	2ΔL	321.51	61.336	293.91	52.496	594.21	23.486	370.73	975.87	198.91	185.32
	K	4	4	4	4	4	4	4	4	4	4
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	p	2.2e-16	1.519e-12	2.2e-16	1.086e-10	2.2e-16	0.0001012	2.2e-16	2.2e-16	2.2e-16	2.2e-16
	χ	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877
M1a vs M2a	lnL	-7228.052639	-1943.028431	-2749.813253	-1919.354366	-6804.813715	-1169.588702	-4009.024521	10796.382824	3363.304373	
	lnL	-7015.381481	-1943.028431	2747.641931	-1919.354367	-6801.142675	-1169.588702	-3981.442258	10767.549294	3363.304385	
	2ΔL	425.34	0	4.3426	-2,00E-06	7.3421	0	55.165	57.667	-2.4e-05	0
	K	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	p	2.2e-16	1	0.114	1	0.02545	1	1.05e-12	3.004e-13	1	1
	χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915
M7 vs M8	lnL	-7260.652141	-1945.359118	-2750.956795	1925.351999	6793.559395	-1171.791179	-4011.459227	10797.123547	3351.699957	
	lnL	-7015.381484	-1942.826664	-2746.445973	-1919.346836	-6785.007090	-1169.662614	-3981.889340	10767.435440	3350.794888	
	2ΔL	490.54	5.0649	9.0216	12.01	17.105	4.2571	59.14	59.376	1.8101	0.55053
	K	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	p	2.2e-16	0.07946	0.01099	0.002466	0.0001931	0.119	1.439e-13	1.278e-13	0.4045	0.7594
	χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915

1001 **Table S19 continuation.** Single-copy orthologous groups (OG) likelihood-rate test summary of the different selection test models

Model	OG31	OG32	OG33	OG34	OG35	OG36	OG37	OG38	OG39	OG40	
M0 vs M3	InL	-8768.449410	-2868.612793	14621.285148	-8894.045372	-3929.296093	-2874.910280	13312.710408	3928.586702	5930.825560	-7009.639226
	InL	-8544.994586	-2687.374379	14061.496172	-8630.576510	-3850.677314	-2853.074629	13065.035639	3710.943836	5773.580811	-6872.342203
	2ΔL	446.91	362.48	1119.6	526.94	157.24	43.671	495.35	435.29	314.49	274.59
	K	4	4	4	4	4	4	4	4	4	4
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	P	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	7.508e-09	2.2e-16	2.2e-16	2.2e-16	2.2e-16
	χ	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877
M1a vs M2a	InL	-8564.647612	-2752.368582	14140.983283	-8644.277513	-3942.697472	-2853.656918	13093.487688	3756.388973	5805.481261	-6890.667418
	InL	-8545.108356	-2690.644641	14061.772045	-8635.257293	-3856.077852	-2853.656918	13093.487688	3712.184796	5805.481261	-6890.667420
	2ΔL	39.079	123.45	158.42	18.04	173.24	0	0	88.408	0	-4,00E-06
	K	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	P	3.267e-09	2.2e-16	2.2e-16	0.0001209	2.2e-16	1	1	2.2e-16	1	1
	χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915
M7 vs M8	InL	-8574.059132	-2753.230060	14141.789697	8646.652568	3947.301727	-2855.538348	13079.471158	3756.977191	5794.915321	6875.501372
	InL	-8547.602397	-2690.644795	14061.493950	-8630.991740	-3856.077856	-2853.096431	13070.395831	3712.245453	5779.817707	-6874.751966
	2ΔL	52.913	125.17	160.59	31.322	182.45	4.8838	18.151	89.463	30.195	1.4988
	K	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	P	3.236e-12	2.2e-16	2.2e-16	1.58e-07	2.2e-16	0.08699	0.0001145	2.2e-16	2.775e-07	0.4726
	χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915

1002 **Table S19 continuation.** Single-copy orthologous groups (OG) likelihood-rate test summary of the different selection test models

Model	OG41	OG42	OG43	OG44	OG45	OG46	OG47	OG48	OG49	OG50	
M0 vs M3	lnL	-1708.733339	-2773.327806	-3118.301126	-3785.486916	-9144.155714	-3347.179895	-3053.244689	1345.203631	6073.114803	-2544.769522
	lnL	-1673.816166	-2741.298992	-3064.991199	-3709.550828	-8887.890899	-3303.396341	-3004.355660	1315.858938	6008.146911	-2494.922373
	2ΔL	69.834	64.058	106.62	151.87	512.53	87.567	97.778	58.689	129.94	99.694
	K	4	4	4	4	4	4	4	4	4	4
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	p	2.46e-14	4.064e-13	2.2e-16	2.2e-16	2.2e-16	2.2e-16	2.2e-16	5.468e-12	2.2e-16	2.2e-16
	χ	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877	9.4877
M1a vs M2a	lnL	-1719.991061	-2753.946889	-3152.415326	-3843.277682	-8919.315424	-3303.994599	-3004.787145	1335.653420	6008.522814	-2505.373631
	lnL	-1676.231683	-2741.944017	-3069.716063	-3713.752060	-8919.315424	-3303.994599	-3004.514350	1315.858938	6008.522814	-2505.373631
	2ΔL	87.519	24.006	165.4	259.05	0	0	0.54559	39.589	0	0
	K	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	p	2.2e-16	6.127e-06	2.2e-16	2.2e-16	1	1	0.7612	2.531e-09	1	1

	χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915
	lnL	-1727.628142	-2767.010532	-3152.738153	-3843.277680	-8900.392501	-3308.237776	-3007.209124	1340.019158	6016.002738	2495.413217
	lnL	-1676.231688	-2741.965899	-3069.716070	-3713.752084	-8891.840868	-3303.653754	-3004.453461	1315.860983	6008.361196	-2495.413624
	2ΔL	102.79	50.089	166.04	259.05	17.103	9.168	5.5113	48.316	15.283	-0.000814
M7 vs M8	K	2	2	2	2	2	2	2	2	2	2
	alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	p	2.2e-16	1.328e-11	2.2e-16	2.2e-16	0.0001932	0.01021	0.06357	3.223e-11	0.0004801	1
	χ	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915	5.9915

1003

1004 **Table S20.** Single-copy orthologous groups (OG) suggested sites to evolve under
 1005 positive and purifying selection based on the site model analysis.

1006

Genes	Number of genes	Sites	M0 vs M3	M1a vs M2a	M7 vs M8
OG01	13	477	M0	M1a	M7
OG02	13	3060	M0	M2a	M7
OG03	13	1371	M0	M1a	M7
OG04	13	450	M0	M1a	M7
OG05	13	738	M0	M1a	M7
OG06	13	2679	M0	M2a	M7
OG07	13	1380	M0	M1a	M7
OG08	13	159	M0	M1a	M7
OG09	13	192	M3	M1a	M7
OG10	13	930	M0	M1a	M7
OG11	13	1617	M0	M1a	M7
OG12	13	1512	M0	-	M7
OG13	13	474	M0	M1a	M7
OG14	13	2178	M0	-	M7
OG15	13	813	M0	M2a	M7
OG16	13	945	M0	-	M7
OG17	13	1296	M0	M2a	M8
OG18	13	1005	M0	-	M7
OG19	13	678	M0	M2a	M7
OG20	13	939	M0	-	M7
OG21	13	1317	M0	M1a	M7
OG22	13	537	M0	-	M7
OG23	13	435	M0	M1a	M7
OG24	13	600	M0	M2a	M7
OG25	13	909	M0	M1a	M7
OG26	13	336	M0	-	M8
OG27	13	606	M0	M1a	M7
OG28	13	1575	M0	M1a	M7
OG29	13	477	M0	M2a	M8
OG30	13	702	M0	-	M8

1007

1008

1009

1010

1011 **Table S20 continuation.** Single-copy orthologous groups (OG) suggested sites to
1012 evolve under positive and purifying selection based on the site model analysis.

1013

Genes	Number of genes	Sites	M0 vs M3	M1a vs M2a	M7 vs M8
OG31	13	1527	M0		M7
OG32	13	435	M0	M1a	M7
OG33	13	1854	M0	M1a	M7
OG34	13	1800	M0	M1a	M7
OG35	13	1029	M0	M1a	M7
OG36	13	879	M0	-	M8
OG37	13	1707	M0	-	M7
OG38	13	480	M0	M1a	M7
OG39	13	840	M0	-	M7
OG40	13	1425	M0	M2a	M8
OG41	13	1425	M0	M1a	M7
OG42	13	609	M0	M1a	M7
OG43	13	795	M0	M1a	M7
OG44	13	912	M0	M1a	M7
OG45	13	1434	M0	-	M7
OG46	13	948	M0	-	M7
OG47	13	696	M0	M2a	M8
OG48	13	297	M0	M1a	M7
OG49	13	1806	M0	-	M7
OG50	13	495	M0	-	M8

1014

1015

1016 **Table S21.** Statistics of the genome assembly using for the REPET pipeline of the 13 aphid and aphid-like species.

1017	Species	AGLY	AGOS	APIS
	Sequences size (nb)	3224	4718	21920
1018	Cum sequence lengths	308065118 bp (308 Mb)	294278726 bp (294 Mb)	541137574 bp (541 Mb)
	mean sequence length	95553 bp	62373 bp	24686 bp
1019	Var	706925616106.20	35438145424.37	2864951237004.97
	Sd	840788.69	188250.22	1692616.68
1020	cv	8.80	3.02	68.56
	n	3224	4718	21920
1021	min	60	889	200
	Q1	1912	1264	812
1022	median	2958	2910	994
	Q3	5141	22426	1557
1023	max	23246776	5574629	170740645
	N50	15	195	2
1024	L50	5864887nt	437960nt	132544852nt
	N75	34	442	3
1025	L75	2348885nt	198105nt	119541763nt
	N90	68	795	285
1026	L90	645927nt	69749nt	43182nt
	GC content	82597793 (26.81%)	82597793 (26.81%)	148699570 (27.48%)
1027	occurrences of A	110206015 (35.77%)	110206015 (35.77%)	175553482 (32.44%)
	occurrences of T	110076105 (35.73%)	110076105 (35.73%)	175485970 (32.43%)
1028	occurrences of G	41308777 (13.41%)	41308777 (13.41%)	74343043 (13.74%)
	occurrences of C	41289016 (13.40%)	41289016 (13.40%)	74356527 (13.74%)
1029	occurrences of N	5185205 (1.68%)	5185205 (1.68%)	41398550 (7.65%)
1030	occurrences of other			2 (0.00%)

1031 **Table S21 continuation.** Statistics of the genome assembly using for the REPET pipeline of the 13 aphid and aphid-like species.

Species	DNOX	DPLA	DVIT
Sequences size (nb)	5637	1457	10492
Cum sequence lengths	395073589 bp (395 Mb)	416559878 bp (416 Mb)	282671353 bp (282 Mb)
mean sequence length	70085 bp	285902 bp	26941 bp
Var	30456114370.67	180777758121.01	10953827094.39
Sd	174516.80	425179.68	104660.53
cv	2.49	1.49	3.88
n	5637	1457	10492
min	928	50051	141
Q1	2242	80625	732
median	10659	146084	1077
Q3	24771	296839	3389
max	2142037	5044119	2080308
N50	281	184	236
L50	397774nt	518248nt	341590nt
N75	661	504	557
L75	169433nt	218156nt	129057
N90	1362	900	1164
L90	37041nt	109054nt	36802
GC content	86035086 (21.78%)	125473212 (30.12%)	75293698 (26.64%)
occurrences of A	104953206 (26.57%)	145476040 (34.92%)	100560015 (35.57%)
occurrences of T	104986705 (26.57%)	145610626 (34.96%)	100538992 (35.57%)
occurrences of G	43027042 (10.89%)	62727755 (15.06%)	37634441 (13.31%)
occurrences of C	43008044 (10.89%)	62745457 (15.06%)	37659257 (13.32%)
occurrences of N	98535224 (24.94%)	0 (0%)	6278648 (2.22%)
occurrences of other	563368 (0.14%)		

1046 **Table S21 continuation.** Statistics of the genome assembly using for the REPET pipeline of the 13 aphid and aphid-like species.

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

1058

1059

1060

Species	ELAN	MCER	MPER
Sequences size (nb)	7146	49286	360
Cum sequence lengths	334868377 bp (334 Mb)	405711039 bp (405 Mb)	395142899 bp (395 Mb)
mean sequence length	46860 bp	8231 bp	1097619 bp
Var	2403432991129.35	208480886.95	79614281131038.03
Sd	1550300.94	14438.87	8922683.52
cv	33.08	1.75	8.13
n	7146	49286	360
min	1000	1001	5000
Q1	1380	1599	9881
median	2195	2957	17275
Q3	3792	7384	32495
max	71231741	265351	105178091
N50	3	4472	3
L50	62861676nt	23273nt	69480500nt
N75	5	12033	4
L75	33516867nt	7629nt	62328371nt
N90	6	25289	6
L90	29687234nt	2853nt	29865500
GC content	84889544 (25.35%)	121142808 (29.86%)	119241258 (30.18%)
occurrences of A	121136053 (36.17%)	142161864 (35.04%)	137745117 (34.86%)
occurrences of T	121083850 (36.16%)	142206642 (35.05%)	137769523 (34.87%)
occurrences of G	42454008 (12.68%)	60572748 (14.93%)	59645355 (15.09%)
occurrences of C	42435536 (12.67%)	60570060 (14.93%)	59595903 (15.08%)
occurrences of N	7758930 (2.32%)	199725 (0.05%)	387001 (0.10%)

1061 **Table S21 continuation.** Statistics of the genome assembly using for the REPET pipeline of the 13 aphid and aphid-like species.

Species	PNIG	RMAI	RPAD	SMIS
Sequences size (nb)	18348	220	2172	653
Cum sequence lengths	375348459 bp (375 Mb)	326023155 bp (326 Mb)	321589008 bp (321 Mb)	397852643 bp (397 Mb)
mean sequence length	20457 bp	1481923 bp	148061 bp	609268 bp
Var	2203736131.51	119546129129072.09	104523010136.11	30160399665805.68
Sd	46943.97	10933715.25	323300.19	5491848.47
cv	2.29	7.38	2.18	9.01
n	18348	220	2172	653
min	1000	1096	1131	2682
Q1	1522	7495	7982	16972
median	3033	20497	19633	29917
Q3	11969	33477	123870	45044
max	631822	94224415	4088110	101470385
N50	1038	2	140	4
L50	103994nt	93298903nt	652723nt	36263045nt
N75	2410	3	321	7
L75	42723nt	76887858nt	298224nt	32651544nt
N90	4986	4	583	9
L90	9877nt	56292413nt	101141nt	28943255nt
GC content	107489442 (28.64%)	90270308 (27.69%)	89439794 (27.81%)	119611961 (30.06%)
occurrences of A	133781559 (35.64%)	117853279 (36.15%)	116094666 (36.10%)	139175655 (34.98%)
occurrences of T	133825498 (35.65%)	117852668 (36.15%)	116054548 (36.09%)	139015827 (34.94%)
occurrences of G	53753361 (14.32%)	45115152 (13.84%)	44701984 (13.90%)	59809085 (15.03%)
occurrences of C	53736081 (14.32%)	45155156 (13.85%)	44737810 (13.91%)	59802876 (15.03%)
occurrences of N	251960 (0.07%)	46900 (0.01%)	0 (0%)	49200 (0.01%)

1062

1063 **Table S22.** Statistics of the transposable elements consensus library from the REPET pipeline of the 13 aphid and aphid-like species.

Species	AGLY	AGOS	APIS
Sequences size (nb)	1161	240	1604
Cum sequence lengths	2933887 bp (2 Mb)	208307 bp	1310621 bp (1 Mb)
mean sequence length	2527 bp	867 bp	817 bp
Var	6686596.36	310162.68	391656.36
Sd	2585.85	556.92	625.82
cv	1.02	0.64	0.77
n	1161	240	1604
min	347	326	331
Q1	718	535	542
median	1483	663	694
Q3	3445	979	938
max	16899	3996	18580
N50	215	68	512
L50	4159nt	945nt	858nt
N75	441	137	963
L75	2545nt	622nt	611nt
N90	712	193	1314
L90	9453nt	500nt	510nt
GC content	803274 (27.38%)	59362 (28.50%)	390102 (29.76%)
occurrences of A	1093377 (37.27%)	74140 (35.59%)	461661 (35.22%)
occurrences of T	1037236 (35.35%)	74805 (35.91%)	458858 (35.01%)
occurrences of G	403085 (13.74%)	29798 (14.30%)	194026 (14.80%)
occurrences of C	400189 (13.64%)	29564 (14.19%)	196076 (14.96%)
occurrences of N	0 (0%)	0 (0%)	0 (0%)

1078 **Table S22 continuation.** Statistics of the transposable elements consensus library from the REPET pipeline of the 13 aphid and aphid-
 1079 like species.

1080

1081

1082

1083

1084

1085

1086

1087

1088

1089

1090

1091

1092

1093

Species	DNOX	DPLA	DVIT
Sequences size (nb)	1262	2965	3535
Cum sequence lengths	893004 bp	6321378 bp (6 Mb)	4620002 bp (4 Mb)
mean sequence length	707 bp	2131 bp	1306 bp
Var	120586.28	5170311.22	1170010.58
Sd	347.26	2273.83	1081.67
cv	0.49	1.07	0.83
n	1262	2965	3535
min	341	327	329
Q1	511	640	566
median	610	1190	916
Q3	794	2746	1602
max	5409	18499	8462
N50	437	515	775
L50	709nt	3869	12759nt
N75	796	1108	1669
L75	551nt	1886nt	971nt
N90	1056	1873	2609
L90	478nt	824nt	572nt
GC content	281641 (31.54%)	1946135 (30.79%)	1417030 (30.67%)
occurrences of A	311582 (34.89%)	2247668 (35.56%)	1634296 (35.37%)
occurrences of T	299605 (33.55%)	2127575 (33.66%)	1568638 (33.95%)
occurrences of G	138254 (15.48%)	975599 (15.43%)	708796 (15.34%)
occurrences of C	143387 (16.06%)	970536 (15.35%)	708234 (15.33%)
occurrences of N	176 (0.02%)	0 (0%)	38 (0.00%)

1094 **Table S22 continuation.** Statistics of the transposable elements consensus library from the REPET pipeline of the 13 aphid and aphid-
 1095 like species.

1096

1097

1098

1099

1100

1101

1102

1103

1104

1105

1106

1107

1108

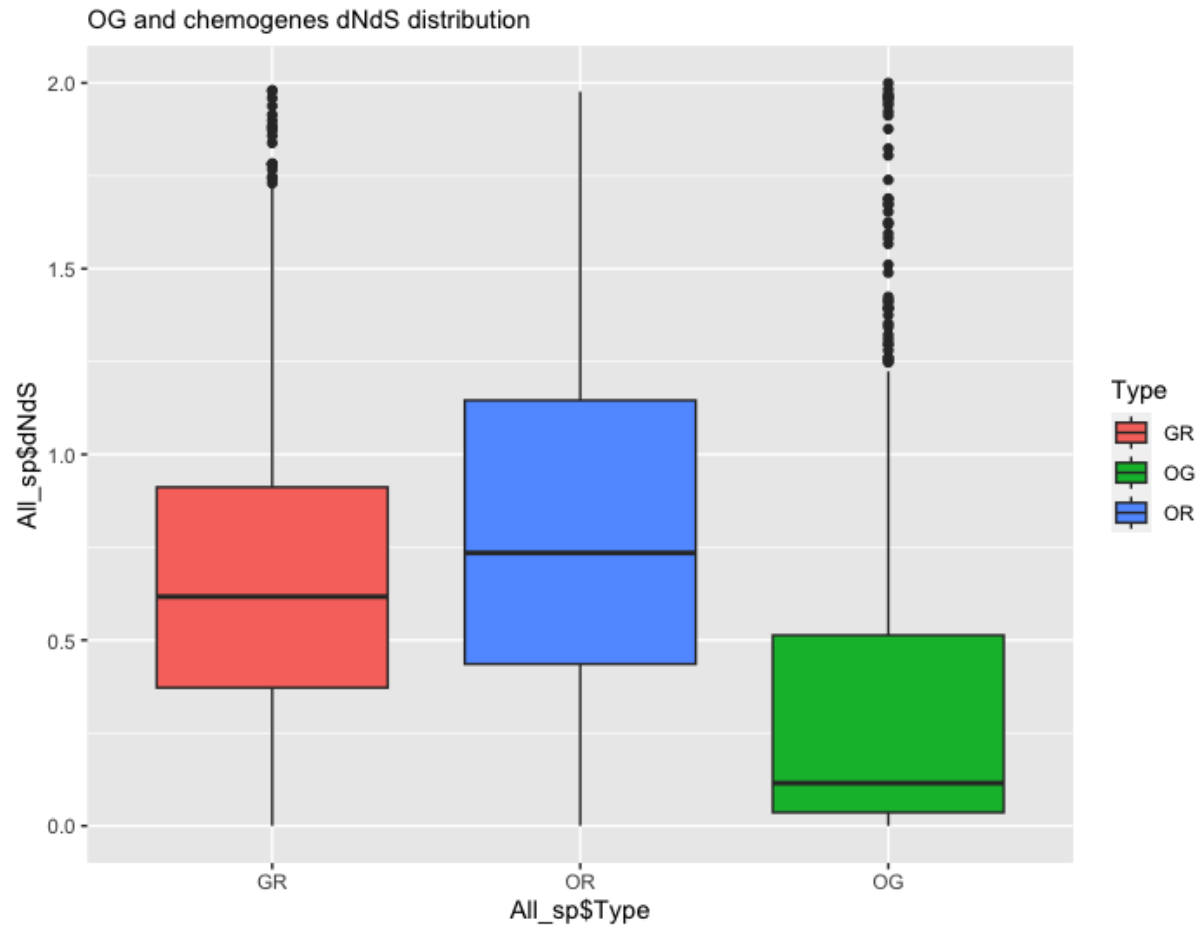
1109

Species	ELAN	MCER	MPER
Sequences size (nb)	1157	1571	1806
Cum sequence lengths	2005004 bp (2 Mb)	1172626 bp (1 Mb)	3758845 bp (3 Mb)
mean sequence length	1732 bp	746 bp	2081 bp
Var	3739808.02	163345.43	5839490.59
Sd	1933.86	404.16	2416.50
cv	1.12	0.54	1.16
n	1157	1571	1806
min	343	332	325
Q1	629	515	626
median	1032	619	1038
Q3	1989	851	2580
max	16846	5297	20355
N50	199	515	281
L50	2612	751nt	4092nt
N75	477	966	632
L75	1273nt	567nt	1691nt
N90	796	1302	1133
L90	701nt	477nt	745nt
GC content	575011 (28.68%)	349195 (29.78%)	1179331 (31.37%)
occurrences of A	735765 (36.70%)	415464 (35.43%)	1326250 (35.28%)
occurrences of T	694168 (34.62%)	407966 (34.79%)	1253264 (33.34%)
occurrences of G	287026 (14.32%)	173180 (14.77%)	586879 (15.61%)
occurrences of C	287985 (14.36%)	176015 (15.01%)	592452 (15.76%)
occurrences of N	60 (0.00%)	1 (0.00%)	0 (0%)

1110 **Table S22 continuation.** Statistics of the transposable elements consensus library from the REPET pipeline of the 13 aphid and
 1111 aphid-like species.

Species	PNIG	RMAI	RPAD	SMIS
Sequences size (nb)	920	830	1018	1516
Cum sequence lengths	1065104 bp (1 Mb)	1597414 bp (1 Mb)	1408656 bp (1 Mb)	2458830 bp (2 Mb)
mean sequence length	1157 bp	1924 bp	1383 bp	1621 bp
Var	786654.68	9038410.91	2499614.73	4049563.55
Sd	886.94	3006.40	1581.02	2012.35
cv	0.77	1.56	1.14	1.24
n	920	830	1018	1516
min	360	340	347	329
Q1	566	629	576	596
median	809	851	764	877
Q3	1400	1633	1451	1649
max	5831	23314	124843	17379
N50	213	85	166	231
L50	1467nt	4206nt	2173nt	2500nt
N75	458	276	434	605
L75	812nt	1212nt	842nt	1128nt
N90	698	551	735	1047
L90	564nt	668nt	591nt	630nt
GC content	343763 (32.28%)	475307 (29.75%)	422297 (29.98%)	774646 (31.50%)
occurrences of A	372134 (34.94%)	567035 (35.50%)	502243 (35.65%)	859947 (34.97%)
occurrences of T	349207 (32.79%)	555072 (34.75%)	484116 (34.37%)	824237 (33.52%)
occurrences of G	166441 (15.63%)	233886 (14.64%)	209464 (14.87%)	380314 (15.47%)
occurrences of C	177322 (16.65%)	241421 (15.11%)	212833 (15.11%)	394332 (16.04%)
occurrences of N	0 (0%)	0 (0%)	0 (0%)	0 (0%)

1112



1113

1114

1115

1116

1117

1118

1119

1120

1121

1122

1123

1124

1125

1126

1127 **Figure S9.** Chemosensory genes (OR and GR) and single-copy orthologous genes dNdS distribution suggest purifying selection.

1128

1129

1130

1131

1132

1133

1134

1135

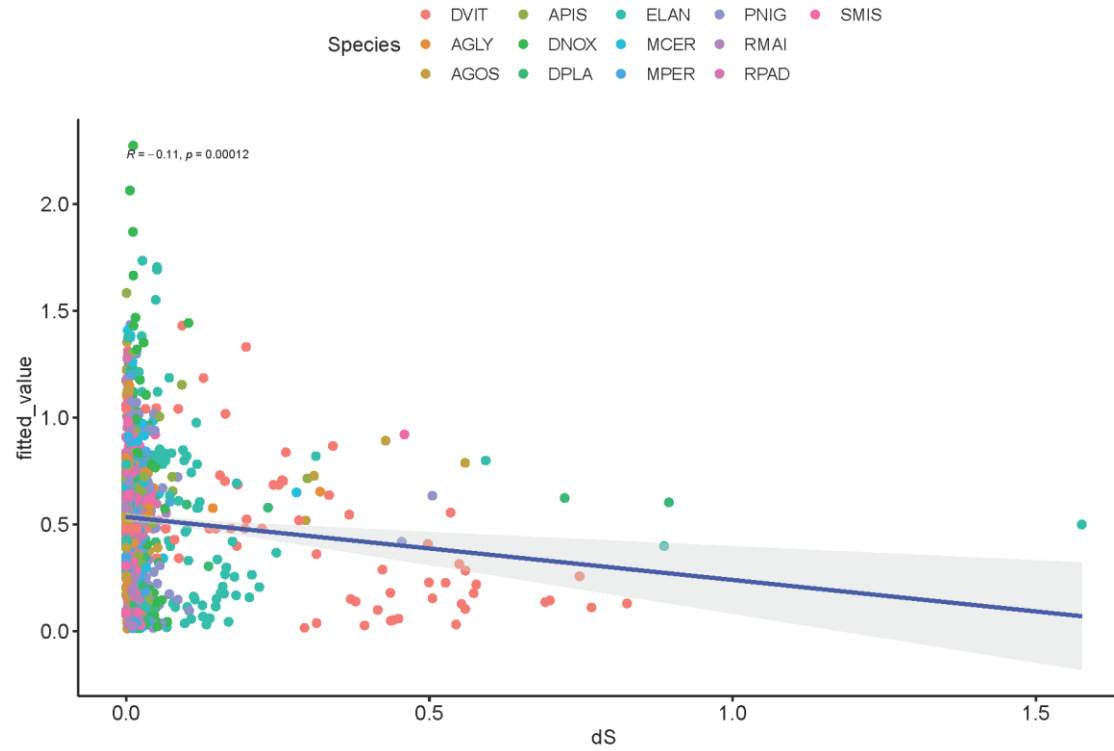
1136

1137

1138

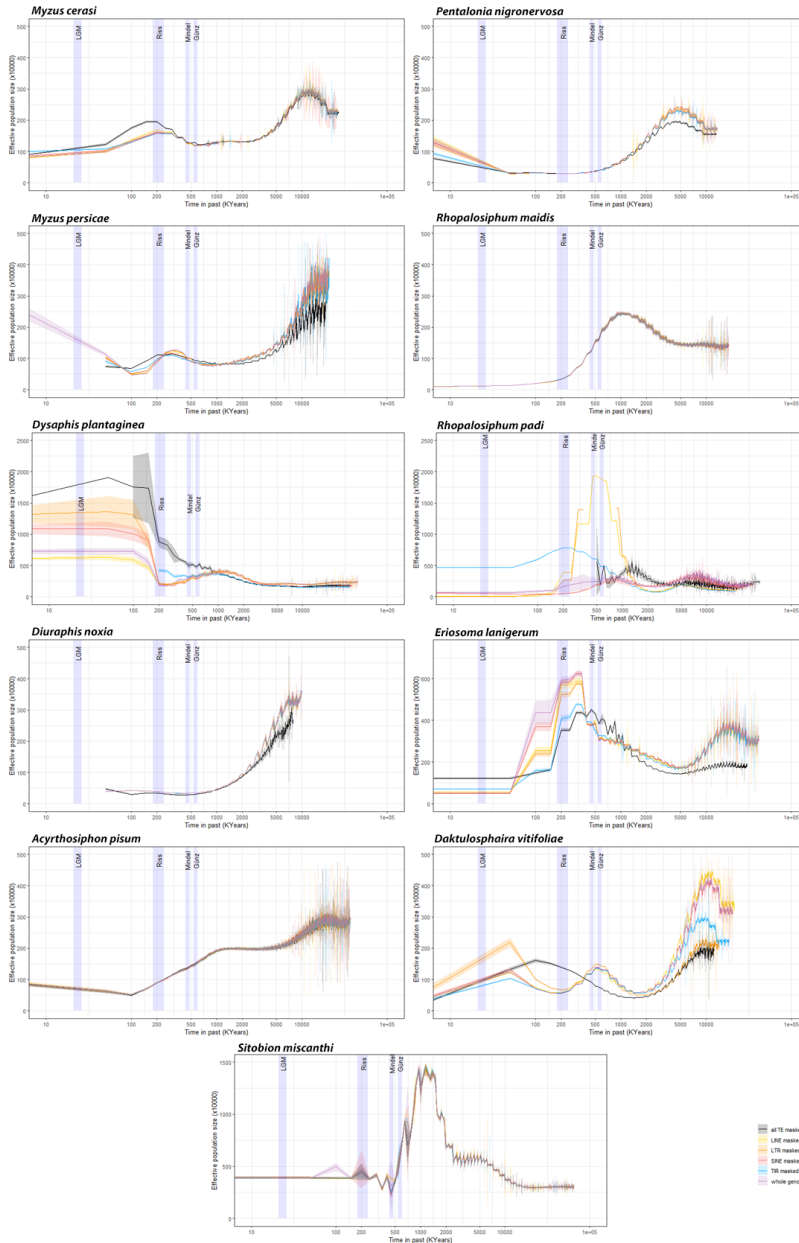
1139

1140



1141 **Figure S10.** Correlation between dN/dS ratio and synonymous divergence (dS) among Gustatory Receptors (GR), Olfactory receptors
1142 (OR), and single copy orthologous groups (OG) for the 13 species.

1143



1146

1147

1148

1149

1150

1151

1152

1153

1154

1155

1156

1157

1158

1159

1160

1161

1162

1163

1164

1165

1166

1167

1168

1169

1170

1171

Figure S11. Changes in effective population size (N_e) estimates from pairwise sequentially Markovian coalescent (PSMC; Li & Durbin, 2011) with unmasked and masked genomes for repeat regions for each aphid species. For *Aphis gossypii*, no short reads data were available and for *Aphis glycines*, the sequencing depth was not sufficient to allow a reliable inference. PSMC curve for *Acyrthosiphon pisum* after masking all of the repeat regions in the genome (black, yellow, LTR, SINE, TIR, one repeat class was unmasked at a time, except the black line which represents all TE class masked) and without masking (purple line). Plots were scaled using one year per generation and mutation rate of 2.7×10^{-10} mutation per site per generation for each aphid species. Ice ages LGM (last glacial maximum: -26.5 to -20 kYears), Riss (-240 to -180 kYears), Mindel (-480 to -430 kYears) and Günz (-600 to -540 kYears) are indicated as grey areas.

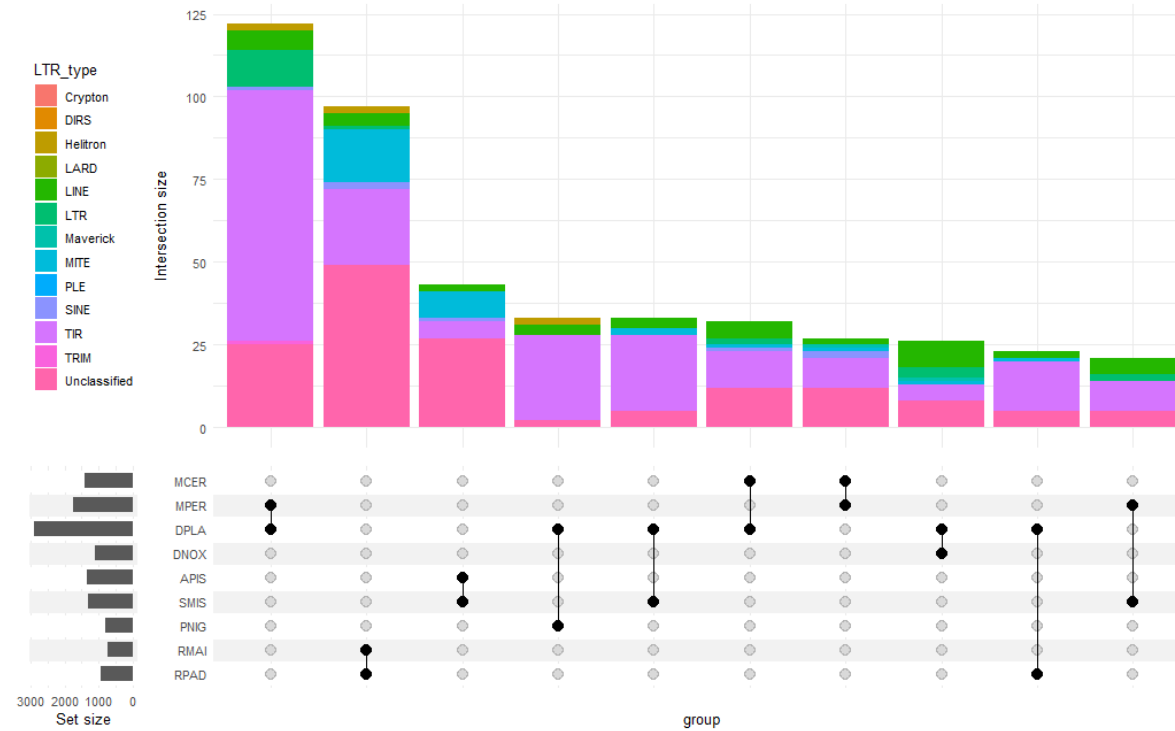


Figure S12. TE-copy identity number among the 12 aphid species using an upset plot (UpSetR R package, Conway et al., 2017).

Table S23. General linear model (GLM) summary of different models to test the distribution of the selection pressure among the different chemosensory genes and the single-copy orthologous gene groups (OG).

Model	VARIABLE	GENES	EFFECT	(ESTIMATE)	DEGREE OF FREEDOM	P-VALUES	FILE
1	dN/dS	OG	Groups	10.987	49	2.2e-16	OG_statistics.txt
2	dN	OG	Groups	1.6818	49	0.003667	OG_statistics.txt
3	dS	OG	Groups	0.4776	49	0.999	OG_statistics.txt
4	dN/dS	OG	Species	1.9703	12	0.02499	OG_spp_statistics.txt
5	dN	OG	Species	114.35	12	2.2e-16	OG_spp_statistics.txt
6	dS	OG	Species	337.72	12	2.2e-16	OG_spp_statistics.txt
7	dN/dS	OG, OR, and GR	Type	119.21	2	2.2e-16	OG_chemo_statistics.txt
8	dN	OG, OR, and GR	Type	12.498	2	4.122e-06	OG_chemo_statistics.txt
9	dS	OG, OR, and GR	Type	8.1958	2	0.0002878	OG_chemo_statistics.txt
10	dN/dS	GR	Groups	3.8772	11	2.509e-05	Genes_GR_dNdS_distribution.txt
11	dN/dS	OR	Groups	3.6779	23	6.749e-08	Genes_OR_dNdS_distribution.txt
12	dN/dS	GR	Species	1.3713	12	0.177	Species_GR_genes_dNdS_distribution.txt
13	dN/dS	OR	Species	0.6161	11	0.8153	Species_OR_genes_dNdS_distribution.txt
14	dN	GR	Groups	0.6372	11	0.7969	Genes_GR_dN_distribution.txt
15	dN	OR	Groups	3.2191	23	1.661e-06	Group_OR_dN_distribution.txt
16	dN	GR	Species	9.0368	12	2.546e-15	Species_GR_genes_dN_distribution.txt
17	dN	OR	Species	3.3247	11	0.0002241	Species_OR_genes_dN_distribution.txt
18	dS	GR	Groups	0.8479	11	0.5921	Genes_GR_dS_distribution.txt
19	dS	OR	Groups	3.9413	23	1.05e-08	Genes_OR_dS_distribution.txt
20	dS	GR	Species	10.508	12	2.2e-16	Species_GR_genes_dS_distribution.txt
21	dS	OR	Species	3.818	11	3.283e-05	Species_OR_genes_dS_distribution.txt

Table S23 continuation. General linear model (GLM) summary of different models to test the distribution of the selection pressure among the different

Model	VARIABLE	GENES	EFFECT	(ESTIMATE)	DEGREE OF FREEDOM	P-VALUES	FILE
22	dN/dS	GR, OR	Chemo (Type)	0.6358	1	0.4255	Chemosensory_genes_dNdS_distribution.txt
23	dN	GR, OR	Chemo (Type)	0.344	1	0.5577	Chemosensory_genes_dN_distribution.txt
24	dS	GR, OR	Chemo (Type)	0.2427	1	0.6224	Chemosensory_genes_dS_distribution.txt
25	dN/dS	GR, OR	Groups	3.8252	35	3.82e-12	Species_chemosensory_genes_dNdS_distribution.txt
26	dN	GR, OR	Groups	1.3409	35	0.09247	Group_chemosensory_genes_dN_distribution.txt
27	dS	GR, OR	Groups	2.0059	35	0.0005995	Group_chemosensory_genes_dS_distribution.txt

Table S24. Locus Overlap Analysis (LOLA) transposable elements (TEs) enrichment analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes in 10Kb length. Excel file:

- a LOLA_summary Overall summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes
- b AGLY_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for Aphis glycine (AGLY) in the Gustatory receptor genes (GR)
- c AGLY_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for Aphis glycine (AGLY) in the Olfactory receptor genes (OR)
- d DPLA_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for Dysaphis plantaginea (DPLA) in the Gustatory receptor genes (GR)
- e DPLA_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for Dysaphis plantaginea (DPLA) in the Olfactory receptor genes (OR)
- f AGOS_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for Aphis gossypii (AGOS) in the Gustatory receptor genes (GR)
- g AGOS_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for Aphis gossypii (AGOS) in the Olfactory receptor genes (OR)
- h APIS_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for Acyrthosiphon pisum (APIS) in the Gustatory receptor genes (GR)
- i APIS_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for Acyrthosiphon pisum (APIS) in the Olfactory receptor genes (OR)

- j DNOX_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Diuraphis noxia* (DNOX) in the Olfactory receptor genes (OR)
- k DNOX_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Diuraphis noxia* (DNOX) in the Gustatory receptor genes (GR)
- l DVIT_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Daktulosphaira vitifoliae* (DVIT) in the Gustatory receptor genes (GR)
- m DVIT_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Daktulosphaira vitifoliae* (DVIT) in the Olfactory receptor genes (OR)
- n ELAN_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Eriosoma lanigerum* (ELAN) in the Gustatory receptor genes (GR)
- o ELAN_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Eriosoma lanigerum* (ELAN) in the Olfactory receptor genes (OR)
- p MCER_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Myzus cerasii* (MCER) in the Gustatory receptor genes (GR)
- q MCER_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Myzus cerasii* (MCER) in the Olfactory receptor genes (OR)
- r PNIG_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Pentalonia nigronervosa* (PNIG) in the Gustatory receptor genes (GR)
- s PNIG_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Pentalonia nigronervosa* (PNIG) in the Olfactory receptor genes (OR)

- t RPAD_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Rhopalosiphum padi* (RPAD) in the Gustatory receptor genes (GR)
- u RPAD_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Rhopalosiphum padi* (RPAD) in the Olfactory receptor genes (OR)
- v RMAI_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Rhopalosiphum maidis* (RMAI) in the Gustatory receptor genes (GR)
- w RMAI_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Rhopalosiphum maidis* (RMAI) in the Olfactory receptor genes (OR)
- x SMIS_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Sitobion miscanthi* (SMIS) in the Gustatory receptor genes (GR)
- y SMIS_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Sitobion miscanthi* (SMIS) in the Olfactory receptor genes (OR)
- z MPER_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Myzus persicae* (MPER) in the Gustatory receptor genes (GR)
- aa MPER_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis for *Myzus persicae* (MPER) in the Olfactory receptor genes (OR)

Table S25. Locus Overlap Analysis (LOLA) transposable elements (TEs) depletion analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes in 10Kb length.. Excel file:

- a AGLY_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for Aphis glycine (AGLY) in the Gustatory receptor genes (GR)
- b AGLY_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for Aphis glycine (AGLY) in the Olfactory receptor genes (OR)
- c DPLA_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for Dysaphis plantaginea (DPLA) in the Gustatory receptor genes (GR)
- d DPLA_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for Dysaphis plantaginea (DPLA) in the Olfactory receptor genes (OR)
- e AGOS_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for Aphis gossypii (AGOS) in the Gustatory receptor genes (GR)
- f AGOS_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for Aphis gossypii (AGOS) in the Olfactory receptor genes (OR)
- g APIS_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for Acyrthosiphon pisum (APIS) in the Gustatory receptor genes (GR)
- h APIS_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for Acyrthosiphon pisum (APIS) in the Olfactory receptor genes (OR)

- i DNOX_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Diuraphis noxia* (DNOX) in the Olfactory receptor genes (OR)
- j DNOX_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Diuraphis noxia* (DNOX) in the Gustatory receptor genes (GR)
- k DVIT_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Daktulosphaira vitifoliae* (DVIT) in the Gustatory receptor genes (GR)
- l DVIT_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Daktulosphaira vitifoliae* (DVIT) in the Olfactory receptor genes (OR)
- m ELAN_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Eriosoma lanigerum* (ELAN) in the Gustatory receptor genes (GR)
- n ELAN_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Eriosoma lanigerum* (ELAN) in the Olfactory receptor genes (OR)
- o MCER_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Myzus cerasii* (MCER) in the Gustatory receptor genes (GR)
- p MCER_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Myzus cerasii* (MCER) in the Olfactory receptor genes (OR)
- q PNIG_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Pentalonia nigronervosa* (PNIG) in the Gustatory receptor genes (GR)
- r PNIG_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Pentalonia nigronervosa* (PNIG) in the Olfactory receptor genes (OR)

- s RMAI_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Rhopalosiphum maidis* (RMAI) in the Gustatory receptor genes (GR)
- t RMAI_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Rhopalosiphum maidis* (RMAI) in the Olfactory receptor genes (OR)
- u RPAD_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Rhopalosiphum padi* (RPAD) in the Gustatory receptor genes (GR)
- v RPAD_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Rhopalosiphum padi* (RPAD) in the Olfactory receptor genes (OR)
- w SMIS_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Sitobion miscanthi* (SMIS) in the Gustatory receptor genes (GR)
- x SMIS_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Sitobion miscanthi* (SMIS) in the Olfactory receptor genes (OR)
- y MPER_GR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Myzus persicae* (MPER) in the Gustatory receptor genes (GR)
- z MPER_OR Summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis for *Myzus persicae* (MPER) in the Olfactory receptor genes (OR)

Table S26. Locus Overlap Analysis (LOLA) transposable elements (TEs) enrichment and depletion analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes in a 2Kb length.

- a LOLA_summary Overall summary of the Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes in 2Kb length.
- b LOLA_summary Locus Overlap Analysis (LOLA) Transposable Elements (TEs) enrichment analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes among the 13 species in 2Kb length.
- c LOLA_summary Locus Overlap Analysis (LOLA) Transposable Elements (TEs) depletion analysis in the Gustatory receptor (GR) and Olfactory receptor (OR) genes among the 13 species in 2Kb length.

5. OUTLINE CHAPTER 2 : COLONIZATION HISTORY OF THE ROSY APPLE APHID

Understanding the evolutionary processes underlying the colonization of new environments and the range expansion of species is a key goal in evolutionary biology (Austerlitz et al., 1997; Excoffier et al., 2009; Rius and Darling, 2014; Hoffman and Courchamp, 2016; Angert et al., 2020). Crop parasites frequently present the colonization of new hosts (host shift) across large geographic ranges and are good models for studying the mechanisms of rapid colonization and range expansion (Stuckenbrock and McDonald, 2008; Gladieux et al., 2014; Garnas et al., 2016). Understanding diversity and population structure is also critical to studying reliable genes involved in adaptation (Tiffin and Ross-Ibarra, 2014), which was the following goal of my Ph.D. project (Chapters 3 and 4).

Using microsatellite markers, we described population structure and diversity and inferred the colonization history of the rosy apple aphid *Dysaphis plantaginea*. The rosy aphid is one of the major pests of the cultivated apple in Europe, North America, and the Middle East. Previous population genetics studies showed that the cultivated apple tree was domesticated in Central Asia, introduced to Europe by the Greeks, and disseminated worldwide by humans (Cornille et al., 2012 and 2014). Nevertheless, there was no information on when and how the colonization of the rosy apple aphid occurred on its primary host, the cultivated apple (Guillemaud et al., 2011), during its domestication. Besides, whether colonization occurred by population contraction followed by expansion, a typical pattern associated with colonization of a new environment, was unknown. Eventually, as aphids are known to be big dispersers (Roderick, 1996), the occurrence of gene flow among aphid populations during colonization was a

central question still unresolved. In addition, endosymbiotic bacteria hosted by aphids are known to play a role in the colonization of new habitats (Tsuchida et al., 2002; Russell et al., 2013; Zepeda-Paulo et al., 2018; Lenhart and White, 2020; Leclair et al., 2021), and in the degree of gene flow among populations (Leonardo and Mondor, 2006). The bacteria community can, therefore, have a role during the colonization of the rosy apple aphid, but we needed to figure it out at the beginning of my Ph.D.

Thus, in this chapter, I present the demographic history of the rosy apple aphid in North Africa, North America, and Europe reconstructed using mitochondrial (*CytB* and *CO1*), bacterial (*16s* rRNA and *TrnpB*), and 30 newly developed microsatellites (or simple sequence repeats, SSR hereafter) markers. Note that the SSR markers were developed upon my arrival. We utilized over 650 rosy apple aphid colonies representing 52 orchards in 13 countries genotyped for the 30 microsatellite markers, a dataset already available before starting my Ph.D. thesis. We evaluated the rosy apple aphid genetic diversity and population structure and characterized the endosymbiont bacteria community diversity. We detected five weakly genetically differentiated rosy apple aphid populations, isolated by distance, with high levels of gene flow. The ABC-RF framework further confirmed the substantial gene flow among populations. We also observed an Eastern-Western decreasing gradient of genetic diversity in Europe, suggesting colonization from Eastern to Western Europe, followed by colonization of North America and Morocco. We tried to infer the population sources and routes of colonization with the ABC-RF framework. However, our power to distinguish among scenarios could have been higher due to the large degree of gene flow among populations and the lack of samples from the Middle East, which is suspected to be a source population. We also found that only the obligatory endosymbiont bacteria *Buchnera aphidicola* Munson was present in *D. plantaginea*

populations across the sampling area, suggesting that secondary endosymbionts did not have a central role during the colonization of the rosy apple aphid.

Our work demonstrated that the recent range expansion of a major fruit tree aphid pest involved high levels of gene flow among genetically diverse populations. The consequences of substantial gene flow within and between populations of the rosy apple aphid remain to be explored with additional genomic markers (*e.g.*, Single Nucleotide Polymorphism or SNPs). Local adaptation results from balancing gene flow and selection (Savolainen et al., 2013; Briscoe et al., 2020). The dependence of aphids on their hosts, as well as certain abiotic factors (Simon et al., 2002; Braendle et al., 2006; Simon et al., 2010), and the recent identification of genes involved in aphid infestation success (Nouhaud et al., 2018, Vazquez et al. 2024 in prep), make aphids a fascinating and relevant model to study the genomic bases of local adaptation. In chapters 3 and 4, I investigate the local adaptation of the rosy apple aphid in the context of recent colonization (Olvera-Vazquez et al., 2021).

Chapter 2, reconstructing the evolutionary history of the rosy apple aphid using SSR markers, has been published in the Peers Community in Evolutionary Biology (<https://doi.org/10.24072/pcjournal.26>). In addition, this work was part of an article supporting science outreach prepared with my supervisor Amandine Cornille and me in the online journal "Passion Entomologie" (<https://passion-entomologie.fr/pommier-puceron-cendre/>).

OUTLINE REFERENCES

Angert, A. L., Bontrager, M. G., & Ågren, J. (2020). What do we really know about adaptation at range edges?. *Annual Review of Ecology, Evolution, and Systematics*, 51, 341-361. <https://doi.org/10.1146/annurev-ecolsys-012120-091002>

Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P. H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical population biology*, 51(2), 148-164. <https://doi.org/10.1006/tpbi.1997.1302>

Braendle, C., Davis, G. K., Brisson, J. A., & Stern, D. L. (2006). Wing dimorphism in aphids. *Heredity*, 97(3), 192-199. <https://doi.org/10.1038/sj.hdy.6800863>

Briscoe Runquist, R. D., Gorton, A. J., Yoder, J. B., Deacon, N. J., Grossman, J. J., Kothari, S., ... & Moeller, D. A. (2020). Context dependence of local adaptation to abiotic and biotic environments: a quantitative and qualitative synthesis. *The American Naturalist*, 195(3), 412-431. <https://doi.org/10.1086/707322>

Cornille, A., Gladieux, P., Smulders, M. J., Roldán-Ruiz, I., Laurens, F., Le Cam, B., ... & Giraud, T. (2012). New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS genetics*, 8(5), e1002703. <https://doi.org/10.1371/journal.pgen.1002703>

Cornille, A., Giraud, T., Smulders, M. J., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, 30(2), 57-65. <https://doi.org/10.1016/j.tig.2013.10.002>

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481–501. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173414>

Garnas, R., J., Auger-Rozenberg, M. A., Roques, A., Bertelsmeier, C., Wingfield, M. J., Saccaggi, D. L., ... & Slippers, B. (2016). Complex patterns of global spread in invasive insects: eco-evolutionary and management consequences. *Biological Invasions*, 18, 935-952. <https://doi.org/10.1007/s10530-016-1082-9>

Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguileta, G., De Vienne, D. M., ... & Giraud, T. (2014). Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular ecology*, 23(4), 753-773. <https://doi.org/10.1111/mec.12631>

Guillemaud, T., Blin, A., Simon, S., Morel, K., & Franck, P. (2011). Weak spatial and temporal population genetic structure in the rosy apple aphid, *Dysaphis plantaginea*, in French apple orchards. *Plos one*, 6(6), e21263.

<https://doi.org/10.1371/journal.pone.0021263>

Hoffmann, B. D., & Courchamp, F. (2016). Biological invasions and natural colonisations: are they that different? *Neobiota*, 29, 1-14. <https://doi.org/10.3897/neobiota.29.6959>

Leclair M, Buchard C, Mahéo F, Simon J-C, Outreman Y (2021) A Link Between Communities of Protective Endosymbionts and Parasitoids of the Pea Aphid Revealed in Unmanipulated Agricultural Systems. *Frontiers in Ecology and Evolution*, 9. <https://doi.org/10.3389/fevo.2021.618331>

Lenhart PA, White JA (2020) Endosymbionts facilitate rapid evolution in a polyphagous herbivore. *Journal of Evolutionary Biology*, 33, 1507–1511. <https://doi.org/10.1111/jeb.13697>

Leonardo, T. E., & Mondor, E. B. (2006). Symbiont modifies host life-history traits that affect gene flow. *Proceedings of the Royal Society B: Biological Sciences*, 273(1590), 1079-1084. <https://doi.org/10.1098/rspb.2005.3408>

Nouhaud, P., Gautier, M., Gouin, A., Jaquiéry, J., Peccoud, J., Legeai, F., ... & Simon, J. C. (2018). Identifying genomic hotspots of differentiation and candidate genes involved in the adaptive divergence of pea aphid host races. *Molecular Ecology*, 27(16), 3287-3300. <https://doi.org/10.1111/mec.14799>

Peace, C. P., Bianco, L., Troggio, M., Van de Weg, E., Howard, N. P., Cornille, A., ... & Vanderzande, S. (2019). Apple whole genome sequences: recent advances and new prospects. *Horticulture Research*, 6. <https://doi.org/10.1038/s41438-019-0141-7>

Olvera-Vazquez, S. G., Alhmedi, A., Miñarro, M., Shykoff, J. A., Marchadier, E., Rousselet, A., ... & Cornille, A. (2021). Experimental test for local adaptation of the rosy apple aphid (*Dysaphis plantaginea*) to its host (*Malus domestica*) and to its climate in Europe. *PCI Ecology*. https://forgemia.inra.fr/amandine.cornille/local_adaptation_dp_ver_4

Olvera-Vazquez ..., Cornille A. (2024) Chemosensory genes involved in host detection evolved by tandem repeat duplications and under strong purifying selection on a large evolutionary timescale. Manuscript in preparation.

Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations?. *Trends in ecology & evolution*, 29(4), 233-242. <https://doi.org/10.1016/j.tree.2014.02.003>

Roderick, G. K. (1996). Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annual review of entomology*, 41(1), 325-352. <https://doi.org/10.1146/annurev.en.41.010196.001545>

Russell JA, Weldon S, Smith AH, Kim KL, Hu Y, Łukasik P, Doll S, Anastopoulos I, Novin M, Oliver KM (2013) Uncovering symbiont-driven genetic diversity across North American pea aphids. *Molecular Ecology*, 22, 2045–2059. <https://doi.org/10.1111/mec.12211>

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11), 807-820. <https://doi.org/10.1038/nrg3522>

Simon, J. C., Rispe, C., & Sunnucks, P. (2002). Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, 17(1), 34-39. [https://doi.org/10.1016/S0169-5347\(01\)02331-X](https://doi.org/10.1016/S0169-5347(01)02331-X)

Simon, J. C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquier, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in functional genomics*, 14(6), 413-423. <https://doi.org/10.1093/bfpg/elv015>

Stukenbrock, E. H., & McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annu. Rev. Phytopathol.*, 46, 75-100. <https://doi.org/10.1146/annurev.phyto.010708.154114>

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12), 673-680. <https://doi.org/10.1016/j.tree.2014.10.004>

Tsuchida T, Koga R, Shibao H, Matsumoto T, Fukatsu T (2002) Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, *Acyrtosiphon pisum*. *Molecular Ecology*, 11, 2123–2135. <https://doi.org/10.1046/j.1365-294X.2002.01606.x>

Zepeda-Paulo F, Ortiz-Martínez S, Silva AX, Lavandero B (2018) Low bacterial community diversity in two introduced aphid pests revealed with 16S rRNA amplicon sequencing. *PeerJ*, 6, e4725. <https://doi.org/10.7717/peerj.4725>

5.1 LARGE-SCALE GEOGRAPHY SURVEY PROVIDES INSIGHTS INTO THE COLONIZATION HISTORY OF A MAJOR APHID PEST ON ITS CULTIVATED APPLE HOST IN EUROPE, NORTH AMERICA AND NORTH AFRICA (FIRST PUBLISHED MANUSCRIPT)

Peer Community Journal

Section: Evolutionary Biology

RESEARCH ARTICLE

Published
2021-11-26

Cite as

S.G. Olvera-Vazquez, C. Remoué, A. Venon, A. Rousselet, O. Grandcolas, M. Azrine, L. Momont, M. Galan, L. Benoit, G. M. David, A. Alhmedi, T. Beliën, G. Alins, P. Franck, A. Haddioui, S.K. Jacobsen, R. Andreev, S. Simon, L. Sigsgaard, E. Guibert, L. Tournant, F. Gazel, K. Mody, Y. Khachtib, A. Roman, T.M. Ursu, I.A. Zakharov, H. Belcram, M. Harry, M. Roth, J.C. Simon, S. Oram, J.M. Ricard, A. Agnello, E. H. Beers, J. Engelman, I. Balti, A. Salhi-Hannachi, H. Zhang, H. Tu, C. Mottet, B. Barrès, A. Degrave, J. Razmjou, T. Giraud, M. Falque, E. Dapena, M. Miñarro, L. Jardillier, P. Deschamps, E. Jousset and A. Cornille (2021) *Large-scale geography survey provides insights into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa*, Peer Community Journal, 1: e34.

Correspondence

amandine.cornille@gmail.com

Peer-review

Peer reviewed and
recommended by

PCI Evolutionary Biology,

<https://doi.org/10.24072/pci.evolbiol.100134>



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

Large-scale geography survey provides insights into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa

S.G. Olvera-Vazquez¹, C. Remoué¹, A. Venon¹, A. Rousselet¹, O. Grandcolas¹, M. Azrine¹, L. Momont¹, M. Galan², L. Benoit², G. M. David³, A. Alhmedi⁴, T. Beliën⁴, G. Alins⁵, P. Franck⁶, A. Haddioui⁷, S.K. Jacobsen⁸, R. Andreev⁹, S. Simon¹⁰, L. Sigsgaard⁸, E. Guibert¹¹, L. Tournant¹², F. Gazel¹³, K. Mody¹⁴, Y. Khachtib⁷, A. Roman¹⁵, T.M. Ursu¹⁵, I.A. Zakharov¹⁶, H. Belcram¹, M. Harry¹⁷, M. Roth¹⁸, J.C. Simon¹⁹, S. Oram²⁰, J.M. Ricard¹¹, A. Agnello²¹, E. H. Beers²², J. Engelman²³, I. Balti²⁴, A. Salhi-Hannachi²⁴, H. Zhang²⁵, H. Tu²⁵, C. Mottet²⁶, B. Barrès²⁶, A. Degrave²⁷, J. Razmjou²⁸, T. Giraud³, M. Falque¹, E. Dapena²⁹, M. Miñarro²⁹, L. Jardillier³, P. Deschamps³, E. Jousset², and A. Cornille¹

Volume 1 (2021), article e34

<https://doi.org/10.24072/pcjournal.26>

Abstract

With frequent host shifts involving the colonization of new hosts across large geographical ranges, crop pests are good models for examining the mechanisms of rapid colonization. The microbial partners of pest insects may also be involved in or affected by colonization processes, which has been little studied so far. We investigated the demographic history of the rosy apple aphid, *Dysaphis plantaginea*, a major pest of the cultivated apple (*Malus domestica*) in Europe, North Africa and North America, as well as the diversity of its microbiota. We genotyped a comprehensive sample of 714 colonies from Europe, Morocco and the US using mitochondrial (CytB and CO1), bacterial (16s rRNA and TrnPB), and 30 microsatellite markers. We detected five populations spread across the US, Morocco, Western and Eastern Europe and Spain. Populations showed weak genetic differentiation and high genetic diversity, except the ones from Morocco and North America that are likely the result of recent colonization events. Coalescent-based inferences revealed high levels of gene flow among populations during the colonization but did not allow determining the sequence of colonization of Europe, North America and Morocco by *D. plantaginea*, likely because of the weak genetic differentiation and the occurrence of gene flow among populations. We found that *D. plantaginea* rarely hosts other endosymbiotic bacteria than its obligate nutritional symbiont *Buchnera aphidicola*. This suggests that secondary endosymbionts did not play an important role in the rapid spread of the rosy apple aphid. These findings have fundamental importance for understanding pest colonization processes and implications for sustainable pest control programs.

¹Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190 Gif-sur-Yvette, France, ²INRAE, UMR 1062, Centre de Biologie pour la Gestion des Populations CBGP (INRAE, IRD, CIRAD, Montpellier SupAgro), Montferrier-sur-Lez, 34980, France, ³Ecologie Systematique et Evolution, CNRS, AgroParisTech, Université Paris-Saclay, 91400, Orsay, France, ⁴Department of Zoology, Fruit Research Center (Pcfrit Npo), Sint-Truiden, Belgium, ⁵Institut de Recerca i Tecnologia Agroalimentàries, IRTA-Fruit Production, PCiTAL, Parc de Gardeny, edifici Fruitcentre, 25003 Lleida, Spain, ⁶INRAE, UR1115, Plantes et Système de cultures Horticoles (PSH), Agroparc CS 40509, Avignon, F-84000, France, ⁷Laboratory of Biotechnology and Valorisation of Plant Genetic Resources, Faculty of Sciences and Techniques, University of Sultan Moulay Slimane, Beni Mellal, Morocco, ⁸Department of Plant and Environmental Sciences, University of Copenhagen, Denmark, ⁹Department of Entomology, Agricultural University, Plovdiv, Bulgaria, ¹⁰Unité Expérimentale UERI 695, INRAE, 460 Chemin de Gothon, F-26320 Saint-Marcel-lès-Valence, France, ¹¹CTIFL, Centre opérationnel de Balandran, Bellegarde, France, ¹²Fédération Régionale de Défense contre les Organismes Nuisibles Nord Pas-de-Calais, Loos-en-Gohelle, France, ¹³Unité Expérimentale Arboricole - UEA - 0393, INRAE, Bordeaux, France, ¹⁴Department of Applied Ecology, Hochschule Geisenheim University, Geisenheim, Germany, ¹⁵NIRDBS, Institute of Biological Research Cluj-Napoca, 400015, Cluj-Napoca, Romania, ¹⁶Vavilov Institute of General Genetics RAS, Moscow, Russia, ¹⁷Laboratoire Evolution, Génomes, Comportement, Écologie, UMR9191 CNRS/IRD/Université Paris-Saclay, Gif-sur-Yvette, France, ¹⁸GAFL, INRAE, 84140, Montfavet, France, ¹⁹INRAE, UMR IGEPP, Domaine de la Motte, Le Rheu, France, ²⁰People's Trust for Endangered Species, London, UK, ²¹Department of Entomology, Cornell AgriTech, Cornell University, Geneva, NY, USA, ²²Department of Entomology, Tree Fruit Research and Extension Center, Washington State University, Wenatchee, WA, USA, ²³Virginia Tech, Alson H. Smith, Jr. Agricultural Research and Extension Center, 595 Laurel Grove Road, Winchester, VA 22602, USA, ²⁴Department of Biology, Faculty of Science of Tunis, University of Tunis El Manar, 2092, Tunis, Tunisia, ²⁵Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, 450009, Zhengzhou, China, ²⁶Université de Lyon, Anses, INRAE, USC CASPER, 69007 Lyon, France, ²⁷IRHS, INRAE, Agrocampus-Ouest, Université d'Angers, SFR 4207 QUASAV, Beaucouzé, France, ²⁸Department of Plant Protection, Faculty of Agriculture and natural Resources, University of Mohaghegh Ardabili, Iran, ²⁹Servicio Regional de Investigación y Desarrollo Agroalimentario, Ctra. AS-267, PK 19, E-33300, Villaviciosa, Asturias, Spain

Peer Community Journal is a member of the
Centre Mersenne for Open Scientific Publishing

<http://www.centre-mersenne.org/>

CENTRE
MERSENNE

Introduction

Understanding the evolutionary processes underlying the colonization of new environments and the range expansion of species is a key goal in evolutionary biology (Austerlitz *et al.* 1997; Excoffier *et al.* 2009; Rius & Darling 2014; Hoffmann & Courchamp 2016; Angert *et al.* 2020). Crop pests in agro-ecosystems, with their frequent colonization of new hosts across large geographic ranges, are good models to study the mechanisms of rapid colonization and range expansion (Stukenbrock & McDonald 2008; Gladieux *et al.* 2014; Garnas *et al.* 2016). The key questions relating to the evolutionary processes underlying the colonization, spread and success of crop pests pertain to the geographic origin of the source population, the location of the colonization routes, the extent to which genetic diversity is reduced via founder effects (Blakeslee *et al.* 2020) and the extent of gene flow among populations during the spread of the plant parasite (Stukenbrock 2016). Current and future threats to biodiversity and their consequences on ecosystem health and services make these questions more relevant than ever. Understanding the routes of pest colonization contributes greatly to the efforts to protect crops against future pest emergence and therefore has direct implications for breeding and agronomic programs that develop biological methods of parasite control (Estoup & Guillemaud 2010; Lawson Handley *et al.* 2011; Turcotte *et al.* 2017; Fraimout *et al.* 2017).

Historically, gene genealogies have been a rich source of information into a species' evolutionary history that could be applied to the study of colonization (Posada & Crandall 2001; Hickerson *et al.* 2010; Bloomquist *et al.* 2010). The characterization of population structure, genetic diversity and demographic history (divergence time, migration rates among populations and effective population size) are also essential to understand the evolutionary processes underlying rapid colonization and range of geographic expansion (Excoffier *et al.* 2009). Approximate Bayesian computation methods (referred to as "ABC" hereafter) provide a robust framework for inferring a species' history by allowing the comparison of alternative demographic models and the estimation of their associated parameters (divergence time, migration rate, effective population size) (Bertorelle *et al.* 2010; Csilléry *et al.* 2010, 2012; Roux & Pannell 2015; Estoup *et al.* 2018; Raynal *et al.* 2019). The power of the ABC methods has made it possible to retrace the evolutionary history of notorious plant parasites (*e.g.*, *Plasmopara viticola* (Berk & Curtis) Berl. & de Toni (Fontaine *et al.* 2021), *Microbotryum lychnidis-dioicae* (DC. Ex Liro) G. Deml & Oberw. (Gladieux *et al.* 2015)), insect crop pests (*e.g.*, *Batrocera dorsalis* Hendel (Aketarawong *et al.* 2014), *Drosophila suzukii* Matsumura (Fraimout *et al.* 2017), *Daktulosphaira vitifoliae* Fitch (Rispe *et al.* 2020) and invasive alien species posing a threat to the native fauna (*e.g.*, *Harmonia axyridis* Pallas (Lawson Handley *et al.* 2011)). These studies have identified the source populations, reconstructed complex colonization routes, and determined the pace of geographic range expansion, often emphasizing the role of human transportation in the spread of these noxious species. Recently, the ABC approach, combined with machine learning (*i.e.*, random forest, referred to as "ABC-RF" hereafter (Estoup *et al.* 2018; Raynal *et al.* 2019), was used to demonstrate that the African arid-adapted locust pest species *Schistocerca gregaria* Forsskål colonized Africa through major migration events driven by the last glacial climatic episodes (Chapuis *et al.* 2020). Yet, ABC methods are underused for estimating the extent of gene flow during parasite colonization (but see (Fraimout *et al.* 2017)). Most studies assume punctual admixture events among populations, but rarely continuous gene flow among populations. Only recently has the new ABC-RF approach been used to infer the invasion routes, evolutionary history and extent of gene flow in the spotted-wing *D. suzukii* Matsumura from microsatellite markers (Fraimout *et al.* 2017) (referred to as SSR for simple sequence repeat hereafter). Beyond population genomics approaches, and in the special case of insect pests, the investigation of colonization history could also benefit from the characterization of insect endosymbiotic bacterial communities. Indeed, many insect pests host a consortium of endosymbiotic bacteria that mediate their adaptation to new environmental conditions (Frago *et al.* 2020). Variations in the endosymbiotic consortium along colonization routes could facilitate the rapid adaptation of insect pests to different environments (Lenhart & White 2020) and, in consequence, its spread. Alternatively, if the colonization stems from only a few populations, it might be accompanied by a loss of endosymbiont diversity along the colonization routes.

Aphids are a good study system to investigate the evolutionary processes involved in range expansion and the colonization of new environments. Aphids infest a wide range of host species and can be major

pests of many crop plants (Blackman & Eastop 2000). Some aphid species have now become cosmopolitan following the dissemination of crops around the globe (Zepeda-Paulo *et al.* 2010; Kirk *et al.* 2013; Brady *et al.* 2014; Zhang *et al.* 2014). The clonal reproduction of aphids during spring and summer is one of the reasons put forward to account for their remarkable success worldwide. Indeed, asexual reproduction allows for a rapid increase in population size after the colonization of a favorable new environment (Simon *et al.* 2002, 2010; Figueroa *et al.* 2018). So far, only a handful of studies have reconstructed the colonization history of aphid crop pests by combining population genetics approaches using the information from SSR, sequence or single nucleotide polymorphism (SNP) markers (Peccoud *et al.* 2008; Zepeda-Paulo *et al.* 2010; Piffaretti *et al.* 2013; Zhang *et al.* 2014; Zhou *et al.* 2015; Kim *et al.* 2016; Fang *et al.* 2018; Morales-Hojas *et al.* 2020; Giordano *et al.* 2020; Leclair *et al.* 2021). These studies demonstrated that aphid species can spread very quickly across the world, probably via plants transported by humans and/or wind. These investigations detected a colonization involving several populations with high genetic diversity, possibly with gene flow (Wei *et al.* 2005; Wang *et al.* 2015) and/or a few “super-clones”, *i.e.*, predominant genotypes widespread in space and time (Vorburger *et al.* 2003; Piffaretti *et al.* 2013; Figueroa *et al.* 2018). Assessing the genetic diversity, genetic structure and the extent of gene flow among populations in aphids are therefore central to determining the evolutionary processes that have occurred during aphid colonization.

Associating the reconstruction of aphid colonization history with the characterization of their endosymbiotic bacterial community can shed light on the processes of their dispersal. Aphids harbor both obligate symbionts that supply them with the nutrients missing from their diet (Buchner 1965) and facultative symbionts that can provide various selective advantages in specific environmental conditions (Haynes *et al.* 2003; Oliver *et al.* 2009). The obligate aphid endosymbiont bacterium *Buchnera aphidicola* has strictly codiverged with most aphid species (Jousselin *et al.* 2009). Bacterial markers (*e.g.*, *TrpB*), along with other typical markers (*e.g.*, *CO1* or *CytB*), can be used to infer the phylogenetic history of aphid species and to investigate signals of recent range expansions (Zhang *et al.* 2014; Popkin *et al.* 2017). Beyond the use of bacterial genomes to help reconstruct aphid phylogeography, the composition of the bacterial populations in aphids might also help to assess the importance of facultative bacteria for colonizing new geographic regions. Many studies have investigated variation in bacterial communities associated with the pea aphid (*Acyrtosiphon pisum* Harris), revealing geographical variation, host-specific differentiation and associations with environmental factors such as temperature, host plants, and natural enemies (Tsuchida *et al.* 2002; Russell *et al.* 2013; Zepeda-Paulo *et al.* 2018; Leclair *et al.* 2021). The effect of endosymbionts on aphid fitness has been confirmed experimentally in certain cases (Leclair *et al.* 2016; Frago *et al.* 2017). Nevertheless, studies on a global scale in non-model aphid species are still scarce (Zytynska & Weisser 2016), and there are as of yet no studies simultaneously investigating the colonization routes of an aphid species and the changes in symbiotic associations along this route.

Dysaphis plantaginea Passerini, the rosy apple aphid, is one of the most harmful aphid pests attacking cultivated apple trees (*Malus domestica* Borkh), causing major economic losses every year, especially in Europe, North Africa and North America (Wilkaniec 1993; Guillemaud *et al.* 2011; Warneys *et al.* 2018). This aphid species occurs across temperate regions (Central and Southwest Asia, North Africa, North America, and Europe) (Blackman & Eastop 2000) in cultivated apple orchards. The rosy apple aphid completes its life cycle on two successive host plants: the cultivated apple trees as its sole primary host plant, from early autumn to late spring, and the plantain herb *Plantago* spp. as a secondary host plant during summer (Bonnemaison 1959). The rosy apple aphid reproduces through cyclical parthenogenesis whereby clonal reproduction alternates with a sexual reproduction, the latter taking place in autumn when females lay fertilized overwintering eggs on apple trees. Eggs hatch in early spring (Blommers *et al.* 2004). While its phylogenetic relationships with other aphid species are quite well resolved, the evolutionary history of *D. plantaginea* has been little explored to date (but see (Guillemaud *et al.* 2011)). The native geographical range of *D. plantaginea* and its ancestral host range are not known. It might have been associated with *Malus sieversii*, the primary ancestor of the cultivated apple (*M. domestica*) (Harris *et al.* 2002), and then colonized Europe during the journey of the cultivated apple along the Silk Routes from Asia to Europe. Alternatively, the rosy apple aphid may have colonized its cultivated apple host in Europe rapidly and recently, about 1,500 years ago when the Greeks brought the cultivated apple to Europe from Central Asia (Cornille *et al.* 2014, 2019). These scenarios are derived from our knowledge of the domestication history of apples; there are no data on the rosy apple aphid that would support any of these

scenarios. The colonization routes of the rosy apple aphid are unknown, except those historical records document the introduction of the rosy apple aphid in North America was recent (ca. 1890s) (Footitt *et al.* 2006). More generally, the population structure and the extent of gene flow among populations throughout the geographic distribution of *D. plantaginea* are largely unknown, including the regions where it causes the most damage in apple orchards, *i.e.*, North America, North Africa and Europe. There are also no data so far regarding the diversity of endosymbionts across a large geographical range in aphids. Here, we investigate the colonization history of this major fruit tree pest using multiple approaches and genetic datasets, drawn from comprehensive samples taken from the primary host, the cultivated apple in Europe, North America and Morocco. Note that, despite repeated attempts, we failed to collect *D. plantaginea* in its putative source region where apple trees originated in Central Asia, preventing us from fully addressing its earliest colonization history. Therefore, we aimed to answer the following questions focusing on regions most negatively affected by the rosy apple aphid, *i.e.*, Europe, North America and North Africa: i) What is the spatial genetic diversity and population structure of *D. plantaginea* across Europe, the US and Morocco? Can we detect genetically differentiated populations, and/or recent bottlenecks in colonizing populations? ii) Did gene flow occur among populations during colonization? iii) Did *D. plantaginea* populations lose or gain symbionts during their colonization?

Material and methods

Samples and DNA extraction

Each sample described hereafter consisted of a single aphid colony of 10-15 females collected on a host plant during the spring of 2017 and 2018. Sampling only one colony per tree ensured that the sample did not contain different clones, which can occur on the same tree. Each colony was kept in ethanol (96%) at -20°C until DNA extraction. For the three methods described below (*i.e.*, SSR genotyping, Sanger sequencing of the aphid mitochondrial *CO1*, *CytB* markers, and *TrpB* bacterial marker, and metabarcoding of the 16S rRNA bacterial marker), DNA was isolated from a single individual per colony using a new standardized protocol (Supplementary material Text S1). We used different individuals to obtain DNA for the amplification of the different markers. For the sake of simplicity, colonies are referred to as 'individual' or 'samples' hereafter. Note that samples used for 16S rRNA sequencing underwent two extra chemical washes before DNA extraction to remove the external bacteria that could be present on the aphid's cuticle. The extra chemical washes consisted of a first wash with dithiothreitol/DTT (50 mM) for 4 minutes, followed by a second wash with potassium hydroxide/KOH (200 mM) for 4 minutes. The KOH wash was performed twice. Five negative controls were also included ((Jousselin *et al.* 2016), Table S1).

Different sample sizes were used for each of the three methods (*i.e.*, SSR genotyping, metabarcoding of the 16S rRNA bacterial marker, Sanger sequencing of the aphid *CO1*, *CytB* and *TrpB* markers). For each sample, the locality, sample collector identity, host plant species, latitude, longitude and use in this study (genotyping, Sanger sequencing and/or metabarcoding) are given in Table S1.

The largest sample was collected for SSR genotyping, comprising 667 *D. plantaginea* samples (colonies) from Europe, Morocco and the US, from three hosts: *M. domestica* (50 sites, *i.e.*, orchards, $N = 654$), but also *M. sylvestris* (one site, Alta Ribagorça in Spain, $N = 7$), the European wild apple, and *P. lanceolata*, the secondary host (one site, Loos-en-Gohelle in France, $N = 6$). The 667 samples originated from 52 different geographic sites (*i.e.*, 52 orchards) spread over 13 countries; seven to 15 individuals were collected at each site (Table S1, Figure S1). We tried to obtain samples from Eastern Asia and Central Asia during fieldwork in 2017 and 2018, and through our collaborative network. However, despite our attempts and although *D. plantaginea* is referenced in the literature on various hosts in several Central Asian countries (Aslan & Karaca 2005; Holman 2009; CAB International 2020), it was not observed in these areas.

For the investigation of the bacterial 16S rRNA region, we used 178 *D. plantaginea* individuals out of the same 667 colonies used for SSR genotyping (Table S1). We selected two to three samples (colonies) per site to cover a wide and even spatial distribution across Europe and North America. The selected 178 individuals were collected across 12 countries on *M. domestica*, except eight on *M. sylvestris* (Table S1). Morocco was the only country not represented for the bacterial 16S rRNA analysis.

For Sanger sequencing of *CO1*, *TrpB* and *CytB*, we used a total of 84 samples belonging to eight aphid species (*D. plantaginea*, *Dysaphis* sp., *Aphis citricola* van der Goot, *Aphis pomi* de Geer, *Aphis spiraeicola* Patch, *Melanaphis pyraria* Passerini, *Myzus persicae* Sulzer, *Rhopalosiphum insertum* Walker) sampled on

five host plant species (*M. domestica*, *M. sylvestris*, *Sorbus aucuparia*, *Prunus persica*, *Pyrus communis*). The species were chosen to represent the aphid genera known to be closely related to the rosy apple aphid (Choi *et al.* 2018). One to three individuals from each of the geographic sites listed above and in Table S1 were sampled, for a total of 67 samples for *D. plantaginea*. The 17 additional samples from seven other aphid species (one to two samples per aphid species) were collected on the cultivated apple and other fruit tree species (Table S1).

PCR and Sanger sequencing

We amplified the coding regions from the aphid mitochondrial cytochrome c oxidase subunit I (*CO1*) gene and the cytochrome B (*CytB*) gene, as well as the Tryptophan synthase subunit B (*TrpB*) from Buchnera. Fragments were amplified following the protocol reported in Popkin *et al.* (2017) with some modifications (Table S2). The final PCR volume was 30 μ L, containing Buffer (1X), MgCl₂ (1.5 mM), dNTP (0.1 mM), Forward and Reverse primers (0.7 μ M), 5 μ L of Taq polymerase (1 U), and 5 μ L of DNA (1/10 dilution). Amplification products were visualized on an agarose gel 1.5% stained with ethidium bromide under ultraviolet light. We prepared four 96-well plates with 15 μ L of PCR products and a negative control. Plates were sent to Eurofins Genomics France SAS for sequencing.

Chromatograms were inspected and corrected manually with CodonCode Aligner version 8.0.1 (www.codoncode.com), assigned as an 'N' when two peaks overlapped. Alignment, evaluation of all coding genes for frameshifts and elimination of pseudogenes were performed with MEGA version 7.0.26 (Kumar *et al.* 2016). The neutral evolution of each gene, and thus its suitability for phylogenetic and population genetic analyses, was assessed with the McDonald and Kreitman test (McDonald & Kreitman 1991; Egea *et al.* 2008) (Table S3). Two samples of *Brachycaudus helichrysi* Kaltenbach, a pest of *Prunus*, for which sequences of *CO1* (NCBI sequence identifiers: KX381827.1, KX381828.1), *CytB* (KX381989.1, KX381990.1), and *TrpB* (KX382153.1, KX38215.1) were available (Popkin *et al.*, 2017), were added in the phylogenetic analyses. Sequences were concatenated, resulting in a data matrix of 86 concatenated sequences.

Phylogenetic tree and taxonomic assignation

We checked the taxonomic assignation of the samples used in this study by running phylogenetic analyses including *D. plantaginea* and other aphid species found on fruit trees with a Bayesian approach implemented in MrBayes v3.2.7 (Huelsenbeck & Ronquist 2001) and Randomized Axelerated Maximum Likelihood-RAxML (Stamatakis 2014), using the GTR "Generalized time-reversible" mutational model. The two *M. pyrarria* individuals were used as an external group. We chose the default parameters (unlink statefreq = (all) revmat = (all) shape = (all); prset applyto = (all) ratepr = variable; mcmc ngen = 1000000 nruns = 2 nchains = 4 samplefreq = 1000 printfreq = 1000). Inferred trees were visualized with FigTree v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

SSR genotyping

We used 30 SSR markers, including one that was previously used for *D. plantaginea* (Guillemaud *et al.* 2011), and 29 that were newly developed from the sequencing of a low coverage genome (see details of the protocol in supplementary material Texts S2 and S3). We tested the neutrality, and the absence of linkage disequilibrium, of the 29 SSR markers using the Ewens-Watterson neutrality test ((Watterson 1978)); Text S3 and Table S4). Each SSR was amplified separately by PCR. PCR was performed in a final volume of 20 μ L (0.2 μ M of each forward and reverse primer, with the forward primer labeled with a fluorescent dye, 0.2 μ M of dNTPs, between 1 and 1.5 mM of MgCl₂, 1X Buffer (5X), 2 μ L of a homemade Taq, 5 μ L of DNA (1/30 dilution) and sterile H₂O to reach the final volume). We used the following PCR program: 94°C for 5 minutes, followed by 35 cycles of 30s at 94°C, 30s at 60 to 65°C and 45s at 72°C, then 5 minutes at 72°C and finally 10 minutes at 4°C. In the PCR program, the annealing temperature varied between 55°C and 65°C. Annealing temperatures for each SSR marker are detailed in Table S5. PCR products were then pooled according to the four multiplexes described in Table S5.

SSR genotyping was performed at the GENTYANE platform (INRAE, Clermont-Ferrand, France). Alleles of each SSR marker were identified, and their size scored with Genemapper v.4.0 (Applied Biosystems TM, Foster City, USA) by two people independently. In case of discrepancy, the electropherogram was triple checked for a final decision. Allele scoring resulting from Genemapper® was then processed with the Autobin Excel Macro (https://www6.bordeaux-aquitaine.inra.fr/biogeco_eng/Scientific-

[Production/Computer-software/Autobin](#)). We retained only multilocus genotypes with less than 30% missing data and containing less than 5% null alleles. Null alleles were detected with GENEPOP v4.7 (Rousset 2008).

Clonal population structure

Each individual was classified according to its multilocus genotype (*MLG*) with GenoDive 2.0b23 ((Meirmans & Tienderen 2004)), Table S1). We used the stepwise mutation model with a threshold of 0 and the corrected Nei's diversity index as a statistic to test clonal population structure. To reduce the influence of clonal copies produced by asexual reproduction on Hardy–Weinberg equilibrium, allelic frequency and genetic differentiation estimates, the dataset was pruned to include only one copy of each *MLG* for further analyses.

Population genetics descriptive statistics

Observed and expected heterozygosity (H_O and H_E) and inbreeding coefficient (F_{IS}) were calculated with GENEPOP v4.7 (Rousset 2008) from the SSR dataset for each SSR marker, each site (*i.e.*, geographic location/orchard) and each population (*i.e.*, clusters inferred with the STRUCTURE software, to comprise individuals with a membership coefficient of at least 62.5% to the given cluster, see results). The 62.5% cut-off was chosen based on the distribution of individual membership coefficients across the clusters detected for the most likely K value (see results). Pairwise genetic differentiation (F_{ST}) between sites and between populations was also calculated with GENEPOP (Rousset 2008). Only sites with at least five successfully genotyped samples were included for the site-specific computations. Allelic richness and private allelic richness for each site and each population were calculated with ADZE (Szpiech *et al.* 2008) using a sample size corresponding to the smallest number of observations per site or population, multiplied by two chromosomes (*e.g.*, a sample size of 20 represents ten individuals x two chromosomes).

We also estimated Nei's nucleotide diversity index π (Nei 1987), Watterson's index q (Watterson 1975), haplotype diversity, Tajima's D (Tajima 1989) and Fu's F_s (Fu 1997) with DNAsp (Rozas *et al.* 2017) using the concatenated three-marker dataset (*i.e.*, *CO1*, *TrpB* and *CytB*) for each population (*i.e.*, cluster inferred with STRUCTURE including individuals with membership coefficient > 62.5% to this cluster, see results).

Detecting recent bottlenecks during population range expansion

We tested whether a bottleneck occurred during the range expansion of each population with the method implemented in BOTTLENECK (Cornuet & Luikart 1996; Piry *et al.* 1999). Inferences regarding historical changes in population size are based on the principle that the expected heterozygosity estimated from allele frequencies decreases faster than the expected heterozygosity estimated under a given mutation model at mutation-drift equilibrium in populations that have experienced a recent reduction in size. The tests were performed under the stepwise-mutation model (SMM) and a two-phase model (TPM) allowing for 30% multi-step changes.

Spatial distribution of allelic richness and observed heterozygosity

Spatial patterns of allelic richness and observed heterozygosity were visualized by mapping the variation in allelic richness and observed heterozygosity at 48 sites in total (*i.e.*, sites with at least five individuals) with the geometry-based inverse distance weighted interpolation in QGIS (Quantum GIS, GRASS, SAGA GIS). The correlation between genetic variability (H_O) and latitude, and between H_O and longitude, was tested using a linear model.

Population subdivision

We inferred the finest population structure by comparing the results obtained with three population genetic tools: STRUCTURE v2.3.2 (Pritchard *et al.* 2000), TESS v2.3.1 (Chen *et al.* 2007), and a discriminant analysis of principal components (DAPC) (Jombart *et al.* 2010). STRUCTURE is based on the use of Markov chain Monte Carlo (MCMC) simulations to infer the assignment of genotypes to K distinct clusters. In addition to this, TESS also considers a spatial component, so that genotypes from sites that are geographically closer to each other are considered more likely to be in the same cluster. For both STRUCTURE (using admixture model with correlated allele frequencies) and TESS (using hierarchical mixture model), ten independent analyses were carried out for each value of K ($1 \leq K \leq 10$ and $2 \leq K \leq 10$,

respectively) with 500,000 MCMC iterations after a burn-in of 50,000 steps. STRUCTURE and TESS outputs were processed with CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) to identify potential distinct modes (*i.e.*, clustering solutions) in replicated runs (10) for each K . We also assessed the population subdivision with DAPC with the R package 'adegenet' (Jombart & Collins 2015), which does not rely on any assumption about the underlying population genetics model, in particular concerning Hardy-Weinberg equilibrium or linkage equilibrium. The number of genetic clusters was investigated with the *find.cluster* function (Jombart & Collins 2015; Ripley & Ripley), which runs successive K -means for clustering. The automatic cluster selection procedure '*diffNgroup*' was used with *n.iter* set to 106 and *n.start* set to 103. The ordination analysis (DAPC) was performed using the *dapc* function. The statistically optimal number of principal components was assessed using the *optim.a.score* function. Assessment of the samples assigned to a genetic cluster was performed using the *compoplot* function.

We used Pophelper (Francis 2016) to run the Evanno method on the STRUCTURE outputs. The Evanno method detects the strongest level of population subdivision (Evanno *et al.* 2005). For TESS, we used the rate of change of the deviation index criterion (*DIC*) to determine the amount of additional information explained by increasing K . For DAPC, we looked at the Bayesian information criterion (*BIC*) obtained with the *adegenet* package to estimate the optimal K value. However, the K identified with the *DIC*, *BIC* and ΔK statistics does often not correspond to the finest biologically relevant population structure (Kalinowski 2011; Cornille *et al.* 2015; Puechmaille 2016). We therefore visualized the bar plots with Pophelper (Francis 2016) and chose the K value for which all clusters had well assigned individuals while no further well-delimited and biogeographically relevant clusters could be identified for higher K values. For further analyses, we considered an individual to be assigned to a cluster when its membership coefficient was $\geq 62.5\%$ to this cluster (see results below).

The spatial pattern of genetic structure was visualized by mapping the mean membership coefficients for each site, as inferred from each of the three population genetics structure analyses, with QGIS 3.12 'Las Palmas' (<https://qgis.org>). We further explored relationships among populations with a principal component analysis performed on a table of standardized alleles frequencies (PCA, *dudi.pca*, *ade4* R package (Dray & Dufour 2007)).

Isolation-by-distance

We tested whether there was a significant isolation-by-distance (IBD) pattern. A Mantel test with 10,000 random permutations was performed between the individual coefficient of relatedness F_{ij} (Loiselle *et al.* 1995) and the matrix of the natural logarithm of geographic distance. We also performed a correlation between $F_{ST}/(1-F_{ST})$ and the natural logarithm of geographic distance. These analyses were performed using SPAGeDI 1.3 (Hardy & Vekemans 2002) separately for each *D. plantaginea* panmictic population (*i.e.*, cluster containing individuals with a membership coefficient > 0.625) identified with TESS, STRUCTURE, and DAPC.

Demographic and divergence history using ABC-RF

We used ABC to investigate whether the spatial patterns of genetic clustering, diversity and differentiation observed in *D. plantaginea* resulted from the occurrence of gene flow among populations during colonization. We also attempted to infer the sequence of colonization events in each population. We used the recently developed ABC method based on a machine learning tool named "random forest" (ABC-RF) to perform model selection and parameter estimations (Pudlo *et al.* 2016; Estoup *et al.* 2018; Raynal *et al.* 2019). In brief, this method creates a "forest" of bootstrapped decision trees to classify scenarios based on the summary statistics of the datasets. Some simulations are not used to build the trees and can thus be used to cross-validate the analysis by computing a prior error rate. This approach allows the comparison of complex demographic models (Pudlo *et al.* 2016) by comparing groups of scenarios with a specific type of evolutionary event with other groups with different types of evolutionary events (instead of considering all scenarios separately) (Estoup *et al.* 2018).

We used a nested ABC approach with two key steps. First, we inferred the divergence and demographic history of the rosy apple aphid in Europe (step 1). Then, we tested the divergence and demographic history of the rosy apple aphid outside of Europe (step 2). Each ABC step compared different sequences of colonization events, with and without bidirectional gene flow among populations (Figure S2). This two-step nested approach avoids the need to compare models that are too complex, which would require the

simulation of too many populations and parameters and is more powerful than testing all scenarios individually to determine the main evolutionary events that characterize demographic history and divergence (Estoup *et al.* 2018). Populations were defined as the clusters detected with STRUCTURE (see results), removing putative admixed individuals (*i.e.*, individuals with a membership coefficient < 0.625 to any given cluster). The model parameters used were the divergence time between X and Y populations (T_{X-Y}), the effective population size of population X (N_{E-X}), the migration rate per generation between X and Y populations (m_{X-Y}). Prior values for divergence time were drawn from the log-uniform distribution bounded between the distributions used in the ABC and are given in Table S6.

For all models, identical SSR datasets were simulated for 29 out of the 30 markers that had perfect repeats (we excluded the L4 marker because it did not have perfect repeats, Tables S4 and S5), increasing confidence in the simulated model. We preliminarily checked that the population structure inferred with 29 SSR markers did not differ significantly from the inferences obtained with 30 SSR markers (data not shown). We assumed a generalized stepwise model of SSR evolution. Mutation rates were allowed to vary across loci, with locus-specific mutation rates drawn from a gamma distribution (α , α/μ) where μ is the mutation rate per generation and α is a shape parameter. We assumed a log-uniform prior distribution for μ ($1e-4$, $1e-3$) and a uniform distribution for α (1.30).

We used ABCtoolbox (Wegmann *et al.* 2010) with fastsimcoal 2.5 (Excoffier & Foll 2011) to simulate datasets, using model parameters drawn from prior distributions (Table S6). We performed 10,000 simulations per scenario how has been suggested (Pudlo *et al.* 2016). For each simulation, we calculated six summary statistics per population with arlsumstats v3.5 (Excoffier & Lischer 2010): H , the mean heterozygosity across loci, $sd(H)$, the standard deviation of the heterozygosity across loci, GW , the mean Garza-Williamson statistic across loci (Garza & Williamson 2001), $sd(GW)$, the standard deviation of the mean Garza-Williamson statistic over populations, NGW , the mean modified Garza-Williamson statistic over loci, $sd(NGW)$, the standard deviation of the mean modified Garza-Williamson statistic over populations. We also computed pairwise F_{ST} (Weir & Cockerham 1984) and genetic distances ($\delta\mu^2$ (Goldstein *et al.* 1995) between pairs of populations.

We used the abcrf v.1.7.0 R statistical package (Pudlo *et al.* 2016) to carry out the ABC-RF analysis. This analysis provides a classification vote that represents the number of times a scenario is selected as the best one among n trees in the constructed random forest. For each ABC step, we selected the scenario, or the group of scenarios, with the highest number of classification votes as the best scenario, or best group of scenarios, among a total of 500 classification trees (Breiman 2001). We then computed the posterior probabilities and prior error rates (*i.e.*, the probability of choosing a wrong group of scenarios when drawing model index and parameter values from the priors of the best scenario) over 10 replicate analyses (Estoup *et al.* 2018). We also checked visually that the simulated models were compatible with the observed dataset by projecting the simulated and the observed datasets onto the two first linear discriminant analysis (LDA) axes (Pudlo *et al.* 2016), and checking that the observed dataset fell within the clouds of simulated datasets.

We calculated parameter inferences using the final selected model following the two-step ABC procedure. Note that the ABC-RF approach includes the model checking step that was performed a posteriori in previous ABC methods.

Characterization of the bacterial community associated with the rosy apple aphid using the 16S rRNA bacterial gene

In order to investigate the bacterial diversity in *D. plantaginea* populations, we amplified a 251 bp portion of the V4 region of the 16S rRNA gene (Mizrahi-Man *et al.* 2013) and used targeted sequencing of indexed bacterial fragments on a MiSeq (Illumina) platform (Kozich *et al.* 2013) following the protocol described in (Jousselin *et al.* 2016). We used 178 aphid DNA extracts (Table S1), comprising 175 *D. plantaginea* individuals and three *M. pyrraria* individuals (Table S1). We wanted to represent a range as large as possible from our samples. We also added eight randomly chosen samples that did not undergo the two extra chemical washes (see Materials and methods, and Table S1). Each sample was amplified twice along with negative controls (DNA extraction and PCR controls). PCR replicates were conducted on distinct 96-well microplates. We obtained a total of 390 PCR products (186 DNA extracts, by 2 for PCR duplicates, plus PCR controls), which were pooled and then separated by gel electrophoresis. Bands based on the expected size of the PCR products were excised from the gel, purified with a PCR clean-up and gel

extraction kit (Macherey-Nagel), and quantified with the Kapa Library Quantification Kit (Kapa Biosystems). Paired-end sequencing of the DNA pool was carried out on a MISEQ (Illumina) FLOWCELL with a 500-cycle Reagent Kit v2 (Illumina).

We first applied sequence filtering criteria following Illumina's quality control procedure. We then used a pre-processing script from (Sow *et al.* 2019) to merge paired sequences into contigs with FLASH V.1.2.11 (Magoč & Salzberg 2011) and trim primers with CUTADAPT v.1.9.1 (Martin 2011). We then used the FROGS pipeline (Escudié *et al.* 2018) to generate an abundance table of bacterial lineages across samples. In brief, we first filtered out sequences > 261 bp and < 241 bp with FROGS, then we clustered variants into operational taxonomic units (OTUs) with SWARM (Mahé *et al.* 2014) using a maximum aggregation distance of three. We identified and removed chimeric variants with VSEARCH (Rognes *et al.* 2016). We only kept OTUs that were present in both PCR replicates of the same sample and then merged the number of reads for each OTU for each aphid sample.

Taxonomic assignment of OTUs was carried out using RDPtools and Blastn (Altschul *et al.* 1990) against the Silva138-16s database (<https://www.arb-silva.de>) as implemented in FROGS. From the abundance table of OTUs across samples, we transformed read numbers per aphid sample into frequencies (percentages); sequences accounting for < 0.5 % of all the reads for a given sample were excluded following (Jousset *et al.* 2016). All filters resulted in excluding reads found in low abundance that could represent sequencing errors and which were also often found in the negative controls.

Results

Taxonomic status of aphid samples

The Bayesian phylogenetic tree built with the 86 sequences representing nine aphid species resulted in a polytomy for the 67 *D. plantaginea* samples from Europe, Morocco and the USA (Figure S3), showing very little sequence variation at the intraspecific level (low bootstrap values < 0.6, Figure S3a). We therefore kept only two representatives out of the 67 *D. plantaginea* individuals in subsequent phylogenetic analyses. Bayesian analysis of this pruned dataset comprising 19 individuals, representing nine aphid species, confirmed known phylogenetic relationships in Aphidinae ((Choi *et al.* 2018), Figure S3b). The species *A. pomi*, *A. citricola* and *A. spiraeicola* were grouped in the same clade, the three species appearing polyphyletic (Figure S3b). *Dysaphis plantaginea* appeared closely related to other representatives of the Macrosiphini tribe (*B. helichrysi* and *M. persicae*). Interestingly, the samples from Iran clustered apart from the *D. plantaginea* clade, the former belonging to the *Dysaphis* sp. species. This result suggests that the two *Dysaphis* sp. Iranian samples belong to a yet unidentified species or population exhibiting strong differentiation from European populations. Those two samples were not included in the population genetics analyses using SSR, as we only had two representatives from Iran (Table S1).

Clone detection

Overall, the proportion of unique genotypes was variable among sites. We found 582 unique genotypes, among which 29 were repeated (12.7% of the total dataset, Table S7) and involved 85 individuals, mainly coming from Belgium (86% of the clones, eight sites; mean proportion of unique genotypes (mean G/N) = 0.36 ± 0.2), Bulgaria (6%, one site, G/N = 0.67), France (5%, three sites, mean G/N = 0.87 ± 0.05), the USA (1 %, one site; G/N = 0.93), Spain (1%, one site; G/N = 0.91) and from the lab-reared aphids (1%, G/N = 0.09). We kept only the 582 unique genotypes for the analyses presented below.

Spatial distribution of allelic variation

The map of allelic richness (Figure 1a) showed that genetic diversity decreased along a northeast to southwest gradient, with the highest allelic richness found in northeastern Europe and the lowest in Morocco and the USA, except for Belgium that showed a lower level of genetic diversity. We found a significant correlation between allelic richness and longitude ($r = 0.229$, P -value = 0.001), as well as between allelic richness and latitude ($r = 0.268$, P -value = 0.001). The map of observed heterozygosity (Figure 1b) confirmed that genetic diversity decreased along a northeast to southwest gradient, with the highest allelic richness in Denmark and the lowest in Morocco. We found a significant correlation between observed heterozygosity and longitude ($r = 0.173$, P -value = 0.003), and between observed heterozygosity and latitude ($r = 0.499$, P -value = $1.95e-08$).

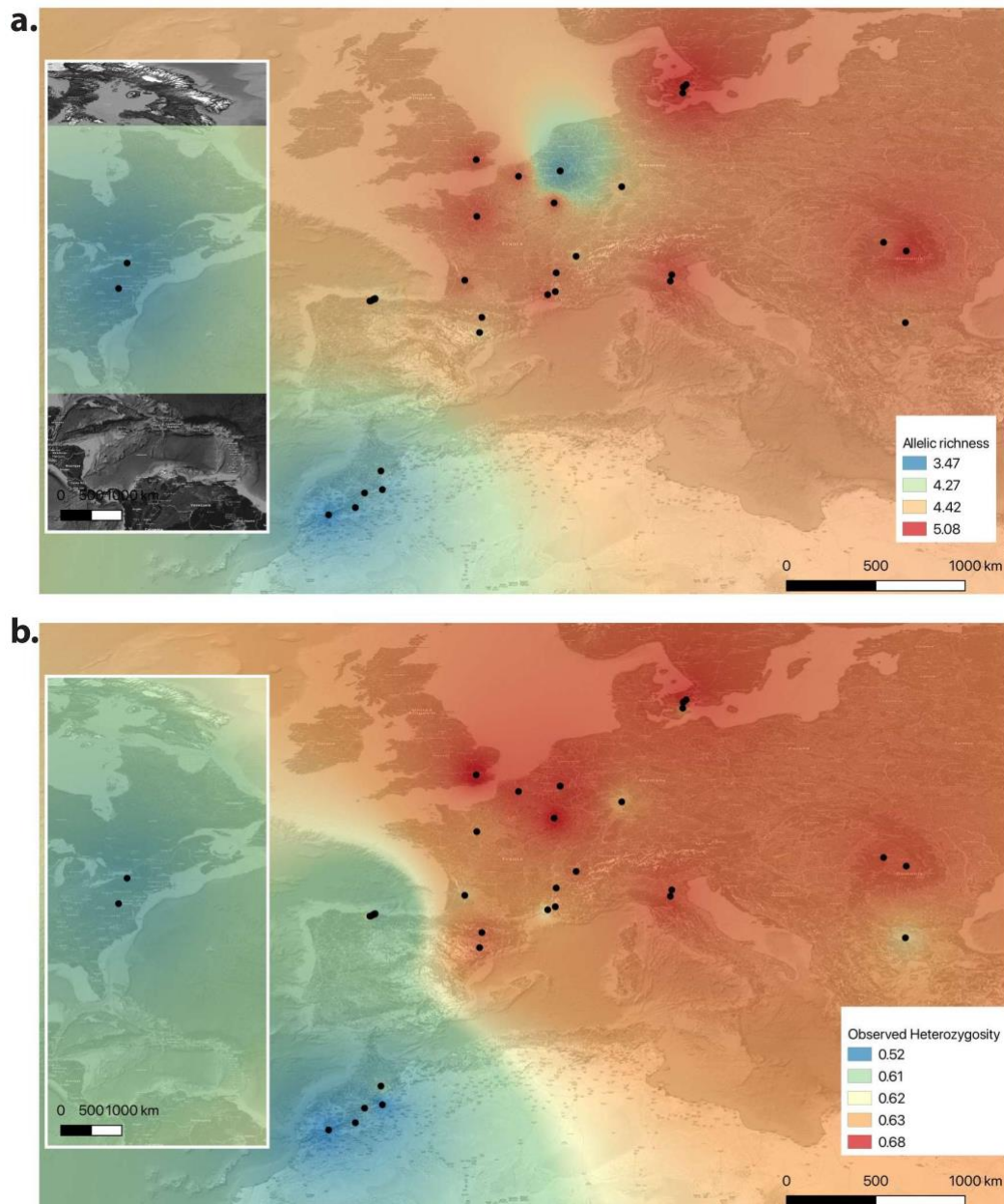


Figure 1. Spatial genetic diversity of *Dysaphis plantaginea* in Europe, Morocco and North America ($N = 582$, clonal copies were excluded, 52 sites, 30 SSR markers). a. Map of allelic richness per site. b. Map of observed heterozygosity. Sites with a sample size below five individuals are not represented on this map. Each dot is a site (*i.e.*, orchard). Red means high allelic richness, blue means low.

Population structure and subdivision

The spatial genetic structures inferred for *D. plantaginea* with TESS, DAPC and STRUCTURE, and the respective DIC , BIC and ΔK , are shown in Figure 2 and supplementary material Figures S4 to S9. For each K value, CLUMPP analyses produced highly similar clustering patterns among the 10 repetitions (average $G > 95\%$). We therefore only presented here the major modes.

With TESS, increasing K above 3 did not reveal any additional cluster (Figure S4). For $K = 3$, TESS analyses revealed a clear partition between the Moroccan samples (blue), and other samples (European and North American, orange). An additional cluster was identified comprising only one Italian individual (yellow) (Figure 2a). With DAPC, increasing K above 4 did not reveal well-delimited new clusters, *i.e.*, only individuals with multiple admixtures were assigned to the new clusters (Figure S5). For $K = 4$, DAPC identified three

well-delimited clusters (Figure 2b): one in Eastern Europe (*i.e.*, Bulgaria, Italy and Romania, in yellow), one in Morocco (blue), one in the USA (orange), plus another cluster that comprised the rest of the Western European (Spain, France, Belgium, the UK) and Danish samples (pink). With STRUCTURE, increasing K above 5 did not reveal additional clusters. For $K = 5$ (Figure 2c), samples were partitioned as follows: Morocco (blue), the USA (orange), Spain (pink), Eastern Europe and Italy (green), and Western Europe and Denmark (yellow).

We used the inferences from STRUCTURE in subsequent analyses because it revealed the finest population genetic structure. Genotypes were then assigned to a given population if their membership coefficient for that population exceeded 0.625 (Tables 1 and S1). We chose this threshold based on the bimodal distribution of cumulative coefficients inferred with STRUCTURE for $K = 5$ (Figure S10). A total of 175 admixed individuals (30% of the dataset) could not be assigned to any population and were not included in subsequent analyses. Most of the admixed individuals were located in Western and Northern Europe; the spatial distribution of the mean number of admixed individuals per site is represented in Figure S11.

We further visualized and quantified the genetic differentiation among populations with a principal component analysis (3D-PCA Figures 2d and S12) and F_{ST} estimates (Figure 2e), respectively. Among all populations, F_{ST} values were low but all significant (Figure 2e). Pairwise genetic differentiation was the lowest between the European populations ($F_{ST} = 0.02$, P -value < 0.0001, Figure 2e); the Moroccan and the USA populations were the most differentiated (Figure 2e). The PCA showed similar genetic relationships among the five populations (Figure 2d).

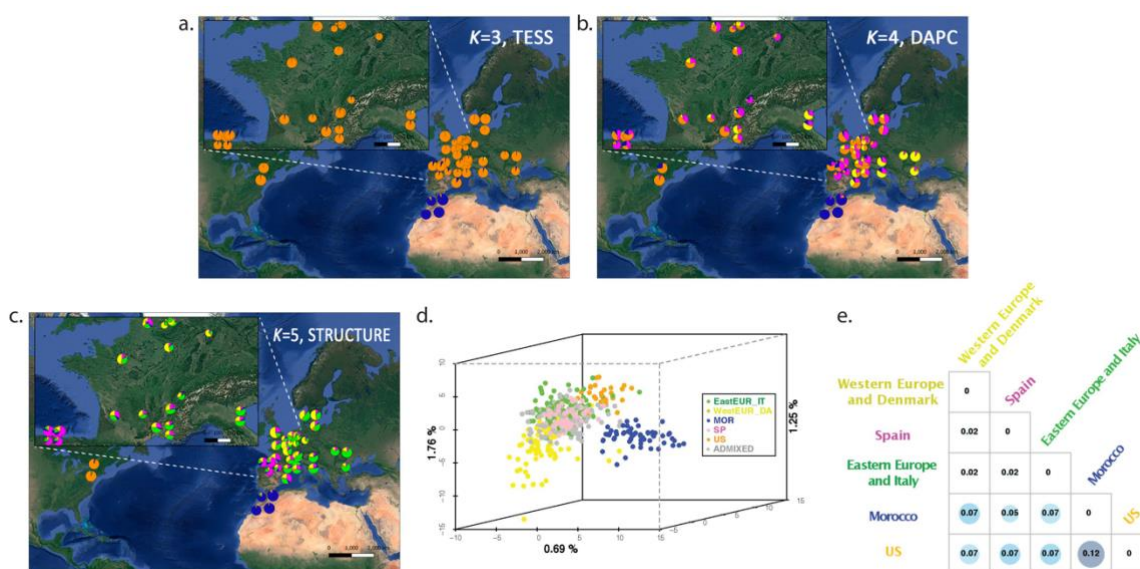


Figure 2. Spatial population structure and differentiation of the rosy apple aphid *Dysaphis plantaginea* in Europe, Morocco and North America ($N = 582$ individuals, 52 sites, 30 SSR markers). Population structure inferred with a. TESS ($K = 3$), b. Discriminant Analysis of Principal Components ($K = 4$), c. STRUCTURE ($K = 5$). Each figure includes a map with each pie chart representing the mean membership coefficients for each site. d. Principal component analysis including the 582 individuals used for running STRUCTURE analyses; each individual is colored according to its membership > 0.625 to one of the five *D. plantaginea* populations detected with STRUCTURE (individuals with a membership coefficient < 62.5% to a given cluster is referred as “admixed”). The size of each pie chart is equivalent to the number of individuals sampled at each site. e. Genetic differentiation (F_{ST}) estimates among the five populations detected with STRUCTURE. The highest F_{ST} values are highlighted by circles (Morocco and the US has the highest value). US: United States of America; MOR: Morocco; SP: Spain; EasternEU_IT: Eastern Europe and Italian population (Bulgaria, Italy, Romania); WestEUR_DA: Western Europe and Danish population (France, Germany, Belgium, United Kingdom, and Denmark).

Genetic diversity, bottleneck and range expansion

The mean expected heterozygosity was relatively high (average = 0.74 with values ranging from 0.55 to 0.75, Table S8). The mean F_{ST} across all loci was low (mean F_{ST} = 0.08, range: -0.006 - 0.5) but significant for all pairs of sites (P -value < 0.001). Allelic richness and private allelic richness were significantly different among the five populations, except between the North American and Moroccan populations for allelic richness, and among the Spanish, Moroccan and Eastern European populations for private allelic richness (Table S9). The North American and Moroccan populations showed significantly lower levels of allelic richness and private allelic diversity than the European populations (*i.e.*, Spanish, Eastern and Western European populations; Tables 1 and S9). In Europe, the Eastern European/Italian population had the highest level of allelic richness and private allelic diversity, followed by the Western European/Danish population, and lastly the Spanish population.

We also tested whether a strong and recent bottleneck occurred for each population. BOTTLENECK analyses showed no significant deviation from the mutation-drift equilibrium in any of the populations (Table S10). Tajima's D and Fu's F_s statistics did not reveal any signature of demographic range expansion either (Table 1).

Isolation-by-Distance (IBD)

A significant but weak IBD pattern was observed for the rosy apple aphid, using the 52 sampling sites ($r^2 = 0.057$, P -value ≤ 0.05), or, excluding the USA sampling sites (*i.e.*, 50 sites, $r^2 = 0.041$, P -value ≤ 0.05), or, using only the European populations (*i.e.*, 45 sites, $r^2 = 0.047$, P -value ≤ 0.05 (Figure S13). The Sp statistic can be used to quantify spatial structure and is useful for comparing populations and/or species. Low Sp values are associated with greater dispersal capacities and/or effective population sizes. Here, Sp values were extremely low (close to 0) and were only significant for the Moroccan and the Western European/Danish populations (Table 1). These results suggest that *D. plantaginea* has high dispersal capacities and/or large effective population sizes.

Inference of the divergence and demographic history of the rosy apple aphid

First, we reconstructed the divergence and demographic history of the rosy apple aphid in Europe (*i.e.*, including only the Spanish, Eastern European/Italian and Western European/Danish populations). We defined 12 scenarios assuming different divergence histories of the Eastern European/Italian, Western European/Danish and Spanish populations (Figure S2). The 12 scenarios were tested with and without gene flow among populations. We therefore ended up comparing 24 scenarios. Classification votes from the first round were the highest ten times out of ten for the group of scenarios that assumed gene flow among the three populations (295 votes out of the 500 RF trees, posterior probability $P = 0.61$, prior error rate = 3.04%, Table S12, Figure S14). Projection of the reference table datasets and the observed dataset on a single axis showed that the observed data fell within the distribution of the simulated summary statistics of the group of scenarios that assumed gene flow among the three populations, suggesting this analysis had the power to discriminate between the two groups of scenarios and to identify the most likely scenario (Figure S14). The second round of ABC inferences testing the sequence of colonization of the European populations requires caution in interpretation as prior error rates were high and posterior probabilities low (Figure S14 and Table S15).

Table 1. Summary statistics obtained for the five *Dysaphis plantaginea* populations (i.e., cluster including individuals with a membership coefficient > 62.5% to a given cluster) inferred with STRUCTURE (K = 5), based on 30 SSR markers and three concatenated genes (*CO1*, *CytB* and *TrpB*, 67 *D. plantaginea* individuals), respectively (left and the right sections of the table).

Population	Statistics inferred from SSR										Statistics inferred from concatenated genes						
	N	H _o	H _E	F _{IS}	A _R (N=28)	A _P (N=28)	Sp	Nseq	NH	HD	S	π	ϕ	Tajima's D	Fu's Fs		
Eastern European and Italian	114	0.68	0.77	0.12 ***	7.77	1.65	0.0006	34	6	0.59	6	0.0005	0.0008	-0.960	0.550		
Western European and Danish	89	0.65	0.74	0.12 ***	7.26	1.17	0.0019*	10	4	0.67	3	0.0006	0.0005	0.120	0.880		
Spanish	113	0.61	0.71	0.14 ***	6.80	0.92	0.0003*	19	5	0.71	4	0.0006	0.0006	-0.001	0.842		
Moroccan	63	0.53	0.61	0.13 ***	5.42	0.60	-0.0038*	2	2	0.67	1	0.0004	0.0003	1.630	1.280		
North American	28	0.57	0.64	0.10 ***	4.70	0.42	-	2	2	0	0	0	0	0	0		

N: number of individuals, H_o and H_E: observed and expected heterozygosity, F_{IS}: inbreeding coefficient, *** P-value < 0.0001, A_R and A_P: allelic richness and private allelic richness, respectively, corrected by the rarefaction method, estimated for a sample size of 28; Sp = spatial parameter; the number of individuals from different geographic locations is too low in the North American population to compute the Sp parameter. Distribution of haplotypes and molecular diversity based on the sequence markers *CO1*, *CytB* and *TrpB* from 67 individuals from different populations based on the genetic structure inferred with STRUCTURE. Nseq = Number of sequences; NH: number of haplotypes; HD: haplotype diversity; S= number of polymorphic sites; π = Nei's nucleotide diversity index; ϕ = Watterson's diversity index ϕ; * = significant value.

We then investigated the colonization history of the rosy apple aphid outside of Europe, *i.e.*, of the Moroccan and North American populations. Given the lack of power to discriminate between different scenarios of colonization sequence of the rosy apple aphid in Europe, and the weak genetic differentiation among the three European populations (mean $F_{ST} = 0.02$, Figure 2e), we merged the three European populations into a single European population ($N = 316$) for this analysis. We then defined six scenarios of sequence of colonization starting either from Europe or Morocco (Figure S2). We excluded the hypothesis that the rosy apple aphid originated in North America. Indeed, historical records show that the introduction of the rosy apple aphid was very recent in America (ca. 1890s) (Footitt *et al.* 2006). Furthermore, the North American population had the lowest levels of private allelic diversity and allelic richness, and the North American samples clustered with the European samples in the DAPC and TESS analyses. For each of the six scenarios, five scenarios of gene flow among populations were tested: no gene flow, gene flow among all populations and gene flow between each population pair (*i.e.*, Europe/Morocco, Europe/the USA, and Morocco/the USA). We simulated these specific models of gene flow among specific pairs of populations for the ABC-RF as we observed variable admixture levels among populations (Figure 2c). In total, 30 scenarios were compared (six colonization sequences x five gene flow modes, Figure S2). ABC-RF analyses showed relatively high support for scenarios assuming gene flow (ABC-RF round 1, 10 out of the 10 replicates for the groups of scenarios assuming gene flow, 337 votes out of the 500 RF trees, posterior probability $P = 0.65$, prior error rate = 6.55%, Table S14, Figure S15). Projection of the reference table datasets and the observed dataset on a single axis showed that the observed data fell within the distribution of the simulated summary statistics of the group assuming gene flow, suggesting the occurrence of gene flow (Figure S15). However, although the observed dataset fell within the distribution of simulated summary statistics (Figure S15), we lacked the power to infer the sequence of colonization of the rosy apple aphid outside of Europe (*i.e.*, posterior probability $P = 0.65$, and high prior error rate = 73.8%, Table S15).

Altogether, ABC-RF inferences supported the occurrence of gene flow during the colonization history of *D. plantaginea*. However, ABC-RF did not allow to determine the sequence of colonization of Europe, North America and Morocco by the rosy apple aphid.

16S rRNA amplicon sequencing

After sequence filtering using the FROGS pipeline, high-throughput sequencing of 16S rRNA bacterial genes from 186 aphids resulted in 5.7 M sequencing reads with an average of 30,800 reads per aphid sample. We found an extremely low bacterial diversity in *D. plantaginea*. The 5.7 M sequencing reads were clustered into 18 OTUs (Figure S16, Table S1). Seven OTUs were assigned to *B. aphidicola* and made up 97.8% of the sequencing reads, 92 % of the reads were assigned to a single *B. aphidicola* OTU, which was found associated with all *D. plantaginea* individuals, the remaining *B. aphidicola* OTUs were found associated with “outgroups”. The remaining reads were mainly assigned to two known aphid endosymbionts, *Serratia symbiotica* and *Regiella insecticola*, which were found in eight and three aphid specimens, respectively. The three aphids hosting *Regiella* belonged to a population of *M. pyrarria* collected in Switzerland, while aphids hosting *Serratia* were found in distantly related populations including *D. plantaginea* collected on *M. domestica* from various apple orchards in France, and *P. communis* in Iran (Table S1). These results highlight the extremely limited diversity of symbionts in the rosy apple aphid across a large geographical scale.

Discussion

We investigated the demographic history of a major fruit tree pest, the rosy apple aphid, in the regions where it impacts the most cultivated apple orchards (*i.e.*, Europe, North Africa and North America). Using multiple approaches, we showed that the colonization of Europe by the rosy apple aphid is likely recent, was not accompanied by strong bottlenecks and involved gene flow between and within populations. The high level of gene flow among populations and within populations was supported by the weak spatial genetic structure observed across Europe, and coalescent-based simulations combined with ABC-RF. We also found that *D. plantaginea* rarely hosts endosymbiotic bacteria other than their primary symbiont, *B. aphidicola* in North America and Europe. Our results provide further understanding of the evolutionary processes at play during pest range expansion.

Colonization of apple trees by the rosy apple aphid is recent, and did not involve drastic bottlenecks

After removing clonal copies from our analyses, we detected with STRUCTURE five main panmictic populations of the rosy apple aphid: three in Europe, one in Morocco and one in the USA. The genetic diversity for each population was within the same range as other aphid species such as *B. helichrysi* (Popkin *et al.* 2017), *Eriosoma lanigerum* (Zhou *et al.* 2015) and *M. persicae nicotianae* (Zepeda-Paulo *et al.* 2010). However, genetic diversity (private alleles and allelic richness) in *D. plantaginea* was higher in Europe and lower in the USA and Morocco. More generally, patterns of genetic structure and diversity suggest different demographic histories for the European, Moroccan and American populations of *D. plantaginea*.

Individuals from the European populations had partial memberships to multiple clusters, with similar membership coefficients for most individuals. This pattern of high admixture along a spatial transect might reflect a continuous gradation in allele frequencies (*i.e.*, a cline) across regions that cannot be detected by the methods used in this study. Indeed, a major limitation of all clustering approaches is the risk of inferring artefactual discrete groups in populations where genetic diversity is distributed continuously. DAPC and STRUCTURE are not immune to this bias and may erroneously identify clusters within a cline (Jombart *et al.* 2010). TESS is more sensitive to allelic gradient; this program includes a decay of the correlation between membership coefficients and distance within clusters (Chen *et al.* 2007). In the presence of clines and with evenly distributed sampling, TESS may detect fewer clusters than STRUCTURE (Durand *et al.* 2009). The larger admixed clusters found with DAPC and STRUCTURE may therefore reflect a cline of allele frequency across Europe for the rosy apple aphid. Allele frequency clines can result from admixture between genetically distinct populations (Menozzi *et al.* 1978; Currat & Excoffier 2005) and/or from subsequent founder events during range expansion (Barbujani *et al.* 1995; Fix 1997; Currat & Excoffier 2005). Genetic diversity is also expected to decrease along the expanding range (Prugnolle *et al.* 2005; François *et al.* 2008). Founder events associated with a recent range expansion may have resulted in the decreasing east-west gradient of genetic diversity and the large number of admixed individuals observed in Europe. It is therefore possible that the European populations of the rosy apple aphid underwent a recent expansion. Note that we did not detect any signature of range expansion with the three-markers dataset (*CO1*, *CytB* and *TrpB*). This lack of signature of range expansion may be due to limited number of samples used in our test (at least for the Moroccan and North American populations), but also that the range expansion is so recent that we cannot catch its footprint with our sequence markers. Several preliminary ABC-RF tests of range expansion of the rosy apple aphid in Europe that failed (*i.e.*, very low posterior probabilities and prior error rates, data not shown) suggest that the second hypothesis is possible.

The Moroccan and North American populations displayed a different pattern of genetic differentiation and diversity compared to the European populations, suggesting an even more recent colonization history. The two populations were well circumscribed and had the highest level of genetic differentiation from the European populations, the lowest genetic diversity and the lowest number of private alleles. In the TESS analysis, the North American samples did not cluster separately from the European samples, and in the DAPC analysis, samples from North America and Western Europe clustered together, suggesting that the North American population originated recently from Europe. This agrees with the earliest record of *D. plantaginea* in the Eastern US dating back to 1890 (Footitt *et al.* 2006). Since then, there have probably been multiple introductions into the USA that prevented genetic differentiation from Europe. Similar scenarios have been described for the tobacco aphid, which was introduced into America from different European gene pools (Zepeda-Paulo *et al.* 2010), and the leaf-curl plum aphid *B. helichrysi* (Piffaretti *et al.* 2013), for which population genetic tools showed very little differentiation between European and North American populations. As for the Moroccan population, we lacked the power to infer its origin with the ABC-RF method. However, the significantly lower genetic diversity and number of private alleles in this population compared with that of Europe, the high level of admixture and the close genetic relationship with the Spanish population, suggest that the Moroccan population resulted from a recent colonization event from Southern Europe. Nevertheless, the history of the rosy apple aphid in North Africa deserves further investigation requiring additional sampling in this region. Altogether, the significantly lower genetic diversity observed in the Moroccan and North American populations suggest that these originated recently through founder events. Indeed, we did not find any evidence that these two populations underwent a recent strong bottleneck despite having significantly lower diversity in both SSR and sequence markers. Thus, the founder effect underlying the colonization of North America and Morocco may involve genetic

drift in small populations rather than severe bottlenecks at the introduction event. Genetic diversity was also found to be higher in native populations within their native range in the tobacco aphid (Zepeda-Paulo *et al.* 2010) and the Russian wheat aphid (Zhou *et al.* 2015).

Altogether, a recent range expansion of the rosy apple aphid on its cultivated apple host is a plausible explanation of the observed spatial genetic structure and diversity. Rapid range expansion has also been described in the Russian wheat aphid (Zhang *et al.* 2014). Unfortunately, the ABC-RF method was not powerful enough to disentangle the different scenarios of colonization of the rosy apple aphid. However, the ABC-RF method was powerful enough to identify the occurrence of gene flow within and outside Europe. Another recent study on the main invasion routes of *D. suzukii* also reported a moderate level of confidence in model choice (*i.e.*, low posterior probabilities between 0.50 to 0.63) and high prior error rates (ranging from 0.30 to 0.40; using ABC-RF (Framout *et al.* 2017)). The absence of samples from key places during the species colonization and individuals of the species ancestral group have been discussed to impact in the resolution of the colonization history analysis using ABC approaches (Lombaert *et al.* 2014). This might account for the lack of support for some of our colonization scenarios using ABC-RF. Additional information on the mutation rates of the newly developed SSR in the present study may also improve support to some of our colonization scenarios. The mutation rate of transcribed and untranscribed SSR was indeed successfully used to reconstruct the main migration routes of *S. gregaria* to Africa using ABC-RF (Chapuis *et al.* 2020).

Colonization with gene flow, likely driven by humans

We found that the expansion of the rosy apple aphid involved several populations with high genetic diversity each, and a high extent of historical gene flow for *D. plantaginea* across Europe, North America and Morocco. Population structure analyses indicated that 30 % of the individuals was a product of recent admixture. Once recently admixed individuals removed, ABC-RF inferences strongly support scenarios with bidirectional gene flow among populations. Both population structure and coalescent-based method inferences therefore indicate the occurrence of recent and ancient gene flow among *D. plantaginea* populations. In addition, the *Sp* parameter estimates revealed large extent of historical gene flow within population. We used the *Sp* parameter estimates to compare the dispersal capacities of *D. plantaginea* with existing estimates in plants (Vekemans & Hardy 2004). The rosy apple aphid showed dispersal capacities equivalent to or even higher than that of wind-dispersed trees. The rosy apple aphid can therefore spread over long distance, as suggested previously (Guillemaud *et al.* 2011), and as frequently found in aphids (Loxdale *et al.* 1993). However, despite its high dispersal capacities and the large amount of gene flow among and within populations, we can still observe a significant (but weak) spatial genetic structure across its distribution. The observed subtle spatial genetic structure may be associated with agricultural practices. The interchange of apple materials (*i.e.*, cultivars in the form of scions and/or trees) is nowadays, and probably has been historically, frequent and a potential way of moving the rosy apple aphid in the form of overwintering eggs among regions where apple is cultivated. This hypothesis agrees with the differentiated population of *D. plantaginea* observed in Asturias (Northwestern Spain). This region is known for the main utilization of the apple for cider production and is based on local cider cultivars (Tardío *et al.* 2021). Most of our samples from Spain are from Asturias, except the Catalan samples that clustered with the European ones. The exchange of apple material within Asturias, and within other regions in Europe, may have occurred, implying higher risk of aphid movement within regions than between regions. Of course, this is not the only explanation of the spatial genetic structure observed, physical barriers (*i.e.*, Atlantic Ocean) could also explain the higher genetic differentiation of the Moroccan and North American populations.

Putative center(s) of origin of the rosy apple aphid

The geographical origin of *D. plantaginea* remains unresolved. Our phylogenetic analyses confirmed previous relationships between the rosy apple aphid and other aphids (Bašilova & Rakauskas 2012; Rebijith *et al.* 2017). However, estimates of the divergence time of *D. plantaginea* from closely related species is now required, with denser sampling of the *Dysaphis* genus across several regions in Eurasia and molecular dating analyses.

Genetic diversity estimates from SSR markers suggest that the source population came from Eastern Europe, but it may also have originated even further east in Central Asia where its fruit tree host was

originally domesticated (Harris *et al.* 2002; Cornille *et al.* 2014). Our failure to collect *D. plantaginea* samples in this area, despite our attempts in China and Kazakhstan, prevents us from testing this scenario. However, note that, while *D. plantaginea* has been recorded in Central Asia according to several faunistic surveys (Kadyrbekov 2002), it is hard to find in these regions. Furthermore, a lack of records of this species in the Global Biodiversity Information Facility (GBIF 2020) suggests *D. plantaginea* is uncommon in this area.

Alternatively, *D. plantaginea* may have originated in the Caucasus or Asia Minor, maybe through a host jump from *Pyrus* to *Malus*. *Dysaphis plantaginea*, *D. radicola*, *D. devecta*, *D. brancoi*, *D. anthrisci*, *D. chaerophylli* are *Dysaphis* species reported to feed on the cultivated apple, *M. domestica*, as their primary host (Blommers *et al.* 2004; Stekolshchikov 2006), but many aphid species also feed on pears, including, *D. reaumuri* Mordvilko and *D. pyri* Boyer de Fonscolombe (Barbagallo *et al.* 2007). The *Pyrus* genus is known to have diverged a long time ago from the genus *Malus* probably in the Caucasus (Celton *et al.* 2009; Xiang *et al.* 2017). Therefore, the ancestral group from which *D. plantaginea* diverged might have had a *Pyrus* species as host. Analysis of samples from these regions is required to test this hypothesis.

Low endosymbiont bacterial diversity associated with *D. plantaginea*

Our results showed that bacterial diversity was strikingly low in *D. plantaginea* across Europe and North America. As expected, *B. aphidicola* was the predominant bacterial species (97% of our reads). We also distinguished different variants of *B. aphidicola* in the Iranian *D. plantaginea* samples which fits with the genetic differentiation observed in the corresponding aphid hosts. At least nine secondary endosymbionts have been reported among aphid species (reviewed in Zytynska & Weisser 2016), including *Arsenophonus Gherma*, *Hamiltonella defensa* Moran, *Regiella insecticola*, *Rickettsia* Da Rocha-Lima, *Rickettsiella Drobne*, *Serratia symbiotica*, *Spiroplasma Saglio* and *Wolbachia* Hertig & Wolbach, and *Fukatsuia symbiotica*. These bacteria have been reported to be aphid facultative symbiont, potential pathogens, or plant-associated microbiota (Gauthier *et al.* 2015). Here, we detected the presence of secondary symbionts only in a few samples. The *Serratia* bacteria was the only secondary symbiont identified and it was found in only eight samples of *D. plantaginea*, collected on *Pyrus communis* in Iran, France and Spain on *M. domestica*. *Regiella* symbiont was found in three samples from another aphid species, *M. pyrararia* collected on *M. domestica*. At least seven endosymbionts (*H. defensa*, *R. insecticola*, *Rickettsia* sp., *Rickettsiella* sp., *S. symbiotica*, *Spiroplasma* sp. and *Wolbachia* sp.) have been reported across a narrower spatial distribution in the model species *Acyrtosiphon pisum* and several other well-studied aphid species (Zytynska & Weisser 2016). By contrast, another study (Henry *et al.* 2015) found that neither *D. plantaginea* nor other *Dysaphis* species hosted any secondary symbionts, although it relied on a small number of individuals. This low endosymbiont bacterial diversity in *D. plantaginea* shows that its likely fast expansion is not the result of an association with different mutualistic endosymbionts.

Concluding remarks

This study has demonstrated that the colonization of a major fruit tree aphid pest occurred without a strong bottleneck, maintaining high genetic diversity, and generated differentiated populations exchanging gene flow, with an isolation-by-distance pattern. The lack of demographic changes in the populations of *D. plantaginea* in Europe, except for the Belgian populations, indicates that seasonal selective pressures, such as insecticide application, have little impact on the genetic diversity of the species. These results may have implications in control and management of *D. plantaginea*, but further studies are needed to fully understand how selective pressures have impacted *D. plantaginea* adaptation. In addition, the use of other genetic markers, such as SNPs, promises to make great strides to elucidate the demographic history of the rosy apple aphid. Finally, the origin of the rosy apple aphid is still unknown. Our results suggest it may have originated in Eastern Europe, the Caucasus or Asia Minor. However, the domestication of the rosy apple aphid primary host (the cultivated apple) in this region remains unknown (Cornille *et al.* 2014, 2019; Spengler 2019). Further investigations on the history of apple domestication, additional sampling of *D. plantaginea* in the Caucasus or Asia Minor, and sampling related aphid species are required to better understand the origin of this major fruit tree pest.

Acknowledgements

We thank the Plateforme de Génotypage GENTYANE INRAE UMR 1095 for assistance in genotyping, and especially the platform leader Charles Poncet. We thank the informatics team at the GQE-Le Moulon, in particular Adrien Falce, Benoit Johannet and Olivier Langella. We are grateful to the Migale bioinformatics Facility (MIGALE, INRAE, 2020, doi: 10.15454/1.5572390655343293E12) for providing help and/or computing and/or storage resources. A previous version of this article has been peer-reviewed and recommended by PCI Evolutionary Biology (<https://doi.org/10.24072/pci.evolbiol.100134>).

Authors' contributions. AC, EJ, PD, LJ conceived and designed the experiments; AC, EJ, PD, LJ, TG, MH obtained funding; AC, AR, AA, TB, GA, FP, AH, SKJ, RA, SS, LS, EG, LT, FG, KM, YK, ARom, TD, IZ, OS, RJM, AA, BEH, EJ, HZ, IB, MR, HT, CM, BB, AD sampled the material; AR, CR, AV, OG, AC, MA, LM, ML, LB, SGOV performed the molecular work; SGOV, AC, OG, MG, LB, EJ, GD analyzed the data. All co-authors discussed the results. The manuscript was written by SGOV, AC, EJ, OG with critical inputs from other co-authors.

Funding

This research was funded by the laboratoire d'Excellence Biodiversity Agrosystem Society and Climate BASC (grant « Emergence POMPUCEDOM ») and Systematic Research funding, ATIP-Avenir program and the Institut Diversité Ecologie et Evolution du Vivant (IDEEV). AR and TU were supported by a grant of the Romanian Ministry of Education and Research, CCCDI - UEFISCDI, project number 384 PED-PN-III-P2-2.1-PED-2019-4924.

Conflict of interest disclosure

The authors declare that they have no financial conflict of interest with the content of this article. Tatiana Giraud is a member of the managing board of PCI Evolutionary Biology. Tatiana Giraud, Amandine Cornille, Emmanuelle Jousselein, and Jean-Christophe Simon are PCI Evolutionary Ecology Biology recommenders.

Data, script and code availability

Data are available online: <https://doi.org/10.5281/zenodo.4537710>

Supplementary information

<https://doi.org/10.5281/zenodo.4537710>

References

- Aketarawong N, Guglielmino CR, Karam N, Falchetto M, Manni M, Scolari F, Gomulski LM, Gasperi G, Malacrida AR (2014) The oriental fruitfly *Bactrocera dorsalis* ss in East Asia: disentangling the different forces promoting the invasion and shaping the genetic make-up of populations. *Genetica*, **142**, 201–213. <https://doi.org/10.1007/s10709-014-9767-4>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology*, **215**, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Angert AL, Bontrager MG, Ågren J (2020) What Do We Really Know About Adaptation at Range Edges? *Annual Review of Ecology, Evolution, and Systematics*, **51**, 341–361. <https://doi.org/10.1146/annurev-ecolsys-012120-091002>
- Aslan B, Karaca İ (2005) Fruit tree aphids and their natural enemies in Isparta region, Turkey. *Journal of pest science*, **78**, 227–229. <https://doi.org/10.1007/s10340-005-0097-2>

- Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H (1997) Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, **51**, 148–164. <https://doi.org/10.1006/tpbi.1997.1302>
- Barbagallo S, Cocuzza G, Cravedi P, Komazaki S (2007) 29 IPM Case Studies: Deciduous Fruit Trees. *Aphids as crop pests*, 651. <https://doi.org/10.1079/9781780647098.0632>
- Barbujani G, Sokal RR, Oden NL (1995) Indo-European origins: A computer-simulation test of five hypotheses. *American Journal of Physical Anthropology*, **96**, 109–132. <https://doi.org/10.1002/ajpa.1330960202>
- Bašilova J, Rakauskas R (2012) Phylogenetic relationships of *Dysaphis pyri* (BOYER DE FONSCOLOMBE) and *Dysaphis reaumuri* (MORDVILKO)(Hemiptera, Sternorrhyncha: Aphididae): COI and EF-1 α evidence. *Organisms diversity & evolution*, **12**, 197–204. <https://doi.org/10.1007/s13127-012-0095-1>
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular ecology*, **19**, 2609–2625. <https://doi.org/10.1111/j.1365-294X.2010.04690.x>
- Blackman RL, Eastop VF (2000) *Aphids on the world's crops: an identification and information guide*. John Wiley & Sons Ltd.
- Blakeslee AMH, Haram LE, Altman I, Kennedy K, Ruiz GM, Miller AW (2020) Founder effects and species introductions: A host versus parasite perspective. *Evolutionary Applications*, **13**, 559–574. <https://doi.org/10.1111/eva.12868>
- Blommers LHM, Helsen HHM, Vaal F (2004) Life history data of the rosy apple aphid *Dysaphis plantaginea* (Pass.)(Homopt., Aphididae) on plantain and as migrant to apple. *Journal of Pest Science*, **77**, 155–163. <https://doi.org/10.1007/s10340-004-0046-5>
- Bloomquist EW, Lemey P, Suchard MA (2010) Three roads diverged? Routes to phylogeographic inference. *Trends in Ecology & Evolution*, **25**, 626–632. <https://doi.org/10.1016/j.tree.2010.08.010>
- Bonnemaison L (1959) Le puceron cendré du pommier (*Dysaphis plantaginea* Pass.) Morphologie et biologie–Méthodes de lutte. *Annales de l'Institut National de la Recherche Agronomique, Série C, Epiphyties*, **3**, 257–322.
- Brady CM, Asplen MK, Desneux N, Heimpel GE, Hopper KR, Linnen CR, Oliver KM, Wulff JA, White JA (2014) Worldwide populations of the aphid *Aphis craccivora* are infected with diverse facultative bacterial symbionts. *Microbial ecology*, **67**, 195–204. <https://doi.org/10.1007/s00248-013-0314-0>
- Breiman L (2001) Random Forests. *Machine Learning*, **45**, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Buchner P (1965) *Endosymbiosis of animals with plant microorganisms*. Interscience Publishers, New York.
- CAB International (2020) *Dysaphis plantaginea*. [Distribution map]. <https://doi.org/10.1079/DMPP20056600429>
- Celton J-M, Chagné D, Tustin SD, Terakami S, Nishitani C, Yamamoto T, Gardiner SE (2009) Update on comparative genome mapping between *Malus* and *Pyrus*. *BMC research notes*, **2**, 182. <https://doi.org/10.1186/1756-0500-2-182>
- Chapuis M-P, Raynal L, Plantamp C, Meynard CN, Blondin L, Marin J-M, Estoup A (2020) A young age of subspecific divergence in the desert locust inferred by ABC random forest. *Molecular Ecology*, **29**, 4542–4558. <https://doi.org/10.1111/mec.15663>
- Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756. <https://doi.org/10.1111/j.1471-8286.2007.01769.x>
- Choi H, Shin S, Jung S, Clarke DJ, Lee S (2018) Molecular phylogeny of Macrosiphini (Hemiptera: Aphididae): An evolutionary hypothesis for the Pterocomma-group habitat adaptation. *Molecular Phylogenetics and Evolution*, **121**, 12–22. <https://doi.org/10.1016/j.ympev.2017.12.021>
- Cornille A, Antolín F, Garcia E, Vernesi C, Fietta A, Brinkkemper O, Kirleis W, Schlumbaum A, Roldán-Ruiz I (2019) A Multifaceted Overview of Apple Tree Domestication. *Trends in Plant Science*, **24**, 770–782. <https://doi.org/10.1016/j.tplants.2019.05.007>
- Cornille A, Feurtey A, Gélín U, Ropars J, Misvanderbrugge K, Gladieux P, Giraud T (2015) Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples. *Evolutionary Applications*, **8**, 373–384. <https://doi.org/10.1111/eva.12250>

- Cornille A, Giraud T, Smulders MJ, Roldán-Ruiz I, Gladieux P (2014) The domestication and evolutionary ecology of apples. *Trends in Genetics*, **30**, 57–65. <https://doi.org/doi:10.1016/j.tig.2013.10.002>
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014. <https://doi.org/10.1093/genetics/144.4.2001>
- Csilléry K, Blum MG, Gaggiotti OE, François O (2010) Approximate Bayesian computation (ABC) in practice. *Trends in ecology & evolution*, **25**, 410–418. <https://doi.org/10.1016/j.tree.2010.04.001>
- Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, **3**, 475–479. <https://doi.org/10.1111/j.2041-210X.2011.00179.x>
- Curat M, Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 679–688. <https://doi.org/10.1098/rspb.2004.2999>
- Dray S, Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software*, **22**, 1–20. <https://doi.org/10.18637/jss.v022.i04>
- Durand E, Chen C, François O (2009) Tess version 2.3—reference manual, August 2009. Available at: memberstimg.imag.fr/Olivier.Francois/tess.html, **1**, 30.
- Egea R, Casillas S, Barbadilla A (2008) Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, **36**, W157–W162. <https://doi.org/10.1093/nar/gkn337>
- Escudé F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet G, Combes S, Pascal G (2018) FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics*, **34**, 1287–1294. <https://doi.org/10.1093/bioinformatics/btx791>
- Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, **19**, 4113–4130. <https://doi.org/10.1111/j.1365-294X.2010.04773.x>
- Estoup A, Raynal L, Verdu P, Marin J-M (2018) Model choice using Approximate Bayesian Computation and Random Forests: analyses based on model grouping to make inferences about the genetic history of Pygmy human populations. *Journal de la Société Française de Statistique*, **159**, 167–190.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Excoffier L, Foll M (2011) fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, **27**, 1332–1334. <https://doi.org/10.1093/bioinformatics/btr124>
- Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173414>
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Fang F, Chen J, Jiang L, Qu Y, Qiao G (2018) Genetic origin and dispersal of the invasive soybean aphid inferred from population genetic analysis and approximate Bayesian computation. *Integrative Zoology*, **13**, 536–552. <https://doi.org/10.1111/1749-4877.12307>
- Figueroa CC, Fuentes-Contreras E, Molina-Montenegro MA, Ramírez CC (2018) Biological and genetic features of introduced aphid populations in agroecosystems. *Current Opinion in Insect Science*, **26**, 63–68. <https://doi.org/10.1016/j.cois.2018.01.004>
- Fix AG (1997) Gene Frequency Clines Produced by Kin-Structured Founder Effects. *Human Biology*, **69**, 663–673. <https://www.jstor.org/stable/41465590>
- Fontaine MC, Labbé F, Dussert Y, Delière L, Richart-Cervera S, Giraud T, Delmotte F (2021) Europe as a bridgehead in the worldwide invasion history of grapevine downy mildew, *Plasmopara viticola*. *Current Biology*. <https://doi.org/10.1016/j.cub.2021.03.009>
- Footitt RG, Halbert SE, Miller GL, Maw E, Russell LM (2006) Adventive aphids (Hemiptera: Aphididae) of America north of Mexico. *Proceedings of the Entomological Society of Washington*, **108**, 583–610 <https://handle.nal.usda.gov/10113/8716>

- Frago E, Mala M, Weldegergis BT, Yang C, McLean A, Godfray HCJ, Gols R, Dicke M (2017) Symbionts protect aphids from parasitic wasps by attenuating herbivore-induced plant volatiles. *Nature Communications*, **8**, 1860. <https://doi.org/10.1038/s41467-017-01935-0>
- Frago E, Zytynska SE, Fatouros NE (2020) Microbial symbionts of herbivorous species across the insect tree. *Mechanisms Underlying Microbial Symbiosis*, 111–159. <https://doi.org/10.1016/bs.aipp.2020.04.002>
- Fraimout A, Debat V, Fellous S, Hufbauer RA, Foucaud J, Pudlo P, Marin J-M, Price DK, Cattel J, Chen X, Deprá M, François Duyck P, Guedot C, Kenis M, Kimura MT, Loeb G, Loiseau A, Martinez-Sañudo I, Pascual M, Polihronakis Richmond M, Shearer P, Singh N, Tamura K, Xuéreb A, Zhang J, Estoup A (2017) Deciphering the Routes of invasion of *Drosophila suzukii* by Means of ABC Random Forest. *Molecular Biology and Evolution*, **34**, 980–996. <https://doi.org/10.1093/molbev/msx050>
- Francis R (2016) POPHELPER: An R package and web app to analyse and visualise population structure. *Molecular Ecology Resources*, **17**, n/a-n/a. <https://doi.org/10.1111/1755-0998.12509>
- François O, Blum MGB, Jakobsson M, Rosenberg NA (2008) Demographic History of European Populations of *Arabidopsis thaliana*. *PLOS Genetics*, **4**, e1000075. <https://doi.org/10.1371/journal.pgen.1000075>
- Fu Y-X (1997) Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics*, **147**, 915.
- Garnas JR, Auger-Rozenberg M-A, Roques A, Bertelsmeier C, Wingfield MJ, Saccaggi DL, Roy HE, Slippers B (2016) Complex patterns of global spread in invasive insects: eco-evolutionary and management consequences. *Biological Invasions*, **18**, 935–952. <https://doi.org/10.1007/s10530-016-1082-9>
- Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318. <https://doi.org/10.1046/j.1365-294X.2001.01190.x>
- Gauthier J-P, Outreman Y, Mieuze L, Simon J-C (2015) Bacterial Communities Associated with Host-Adapted Populations of Pea Aphids Revealed by Deep Sequencing of 16S Ribosomal DNA. *PLOS ONE*, **10**, e0120664. <https://doi.org/10.1371/journal.pone.0120664>
- GBIF (2020) *Dysaphis plantaginea* (Passerini, 1860) in GBIF Secretariat (2021). GBIF Backbone Taxonomy. Checklist dataset <https://doi.org/10.15468/39omei> accessed via GBIF.org on 2021-10-29.
- Giordano R, Donthu RK, Zimin AV, Julca Chavez IC, Gabaldon T, van Munster M, Hon L, Hall R, Badger JH, Nguyen M, Flores A, Potter B, Giray T, Soto-Adames FN, Weber E, Marcelino JAP, Fields CJ, Voegtlin DJ, Hill CB, Hartman GL, Akraiko T, Aschwanden A, Avalos A, Band M, Bonning B, Bretaudeau A, Chiesa O, Chirumamilla A, Coates BS, Cocuzza G, Cullen E, Desborough P, Diers B, DiFonzo C, Heimpel GE, Herman T, Huang Y, Knodel J, Ko C-C, Labrie G, Lagos-Kutz D, Lee J-H, Lee S, Legeai F, Mandrioli M, Manicardi GC, Mazzoni E, Melchiori G, Micijevic A, Miller N, Nasuddin A, Nault BA, O'Neal ME, Panini M, Pessino M, Prischmann-Voldseth D, Robertson HM, Sijun Liu, Song H, Tilmon K, Tooker J, Wu K, Zhan S (2020) Soybean aphid biotype 1 genome: Insights into the invasive biology and adaptive evolution of a major agricultural pest. *Insect Biochemistry and Molecular Biology*, **120**, 103334. <https://doi.org/10.1016/j.ibmb.2020.103334>
- Gladieux P, Feurtey A, Hood ME, Snirc A, Clavel J, Dutech C, Roy M, Giraud T (2015) The population biology of fungal invasions. *Molecular Ecology*, **24**, 1969–1986. <https://doi.org/10.1111/mec.13028>
- Gladieux P, Ropars J, Badouin H, Branca A, Aguilera G, De Vienne DM, Rodríguez de la Vega RC, Branco S, Giraud T (2014) Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular ecology*, **23**, 753–773. <https://doi.org/10.1111/mec.12631>
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463.
- Guillemaud T, Blin A, Simon S, Morel K, Franck P (2011) Weak Spatial and Temporal Population Genetic Structure in the Rosy Apple Aphid, *Dysaphis plantaginea*, in French Apple Orchards. *PLOS ONE*, **6**, e21263. <https://doi.org/10.1371/journal.pone.0021263>
- Hardy OJ, Vekemans X (2002) spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620. <https://doi.org/10.1046/j.1471-8286.2002.00305.x>
- Harris SA, Robinson JP, Juniper BE (2002) Genetic clues to the origin of the apple. *TRENDS in Genetics*, **18**, 426–430. [https://doi.org/10.1016/S0168-9525\(02\)02689-6](https://doi.org/10.1016/S0168-9525(02)02689-6)
- Haynes S, Darby AC, Daniell TJ, Webster G, van Veen FJF, Godfray HCJ, Prosser JI, Douglas AE (2003) Diversity of Bacteria Associated with Natural Aphid Populations. *Applied and Environmental Microbiology*, **69**, 7216–7223. <https://doi.org/10.1128/AEM.69.12.7216-7223.2003>

- Henry LM, Maiden MCJ, Ferrari J, Godfray HCJ (2015) Insect life history and the evolution of bacterial mutualism. *Ecology Letters*, **18**, 516–525. <https://doi.org/10.1111/ele.12425>
- Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010) Phylogeography's past, present, and future: 10 years after. *Molecular phylogenetics and evolution*, **54**, 291–301. <https://doi.org/10.1016/j.ympev.2009.09.016>
- Hoffmann BD, Courchamp F (2016) Biological invasions and natural colonisations: are they that different? *NeoBiota*, **29**, 1–14. <https://doi.org/10.3897/neobiota.29.6959>
- Holman J (2009) *Host plant catalog of aphids*. Springer. <https://doi.org/10.1007/978-1-4020-8286-3>
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Jombart T, Collins C (2015) A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0. *London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling*.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94. <https://doi.org/10.1186/1471-2156-11-94>
- Jousselin E, Clamens A-L, Galan M, Bernard M, Maman S, Gschloessl B, Duport G, Meseguer AS, Clevro F, Coeur d'acier A (2016) Assessment of a 16S rRNA amplicon Illumina sequencing procedure for studying the microbiome of a symbiont-rich aphid genus. *Molecular Ecology Resources*, **16**, 628–640. <https://doi.org/10.1111/1755-0998.12478>
- Jousselin E, Desdevises Y, Coeur d'acier A (2009) Fine-scale cospeciation between *Brachycaudus* and *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in aphids. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 187–196. <https://doi.org/10.1098/rspb.2008.0679>
- Kadyrbekov R (2002) About the aphid fauna (Homoptera Aphididae) of West Tien-Shan (Kazakhstan part). *Tethys Entomological Research*, **4**, 65–76.
- Kalinowski ST (2011) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity*, **106**, 625–632. <https://doi.org/10.1038/hdy.2010.95>
- Kim H, Hoelmer K, Lee S (2016) Population genetics of the soybean aphid in North America and East Asia: test for introduction between native and introduced populations. *Biological Invasions*, **19**. <https://doi.org/10.1007/s10530-016-1299-7>
- Kirk H, Dorn S, Mazzi D (2013) Molecular genetics and genomics generate new insights into invertebrate pest invasions. *Evolutionary Applications*, **6**, 842–856. <https://doi.org/10.1111/eva.12071>
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*, **79**, 5112–5120. <https://doi.org/10.1128/AEM.01043-13>
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, **33**, 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Lawson Handley L-J, Estoup A, Evans DM, Thomas CE, Lombaert E, Facon B, Aebi A, Roy HE (2011) Ecological genetics of invasive alien species. *BioControl*, **56**, 409–428. <https://doi.org/10.1007/s10526-011-9386-2>
- Leclair M, Buchard C, Mahéo F, Simon J-C, Outreman Y (2021) A Link Between Communities of Protective Endosymbionts and Parasitoids of the Pea Aphid Revealed in Unmanipulated Agricultural Systems. *Frontiers in Ecology and Evolution*, **9**. <https://doi.org/10.3389/fevo.2021.618331>
- Leclair M, Pons I, Mahéo F, Morlière S, Simon J-C, Outreman Y (2016) Diversity in symbiont consortia in the pea aphid complex is associated with large phenotypic variation in the insect host. *Evolutionary Ecology*, **30**, 925–941. <https://doi.org/10.1007/s10682-016-9856-1>
- Lenhart PA, White JA (2020) Endosymbionts facilitate rapid evolution in a polyphagous herbivore. *Journal of Evolutionary Biology*, **33**, 1507–1511. <https://doi.org/10.1111/jeb.13697>

- Loiselle B, Sork V, Nason J, Graham C (1995) Loiselle BA, Sork VL, Nason J, Graham C.. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* 82: 1420-1425. *American Journal of Botany*, **82**, 1420–1425. <https://doi.org/10.2307/2445869>
- Lombaert E, Guillemaud T, Lundgren J, Koch R, Facon B, Grez A, Loomans A, Malausa T, Nedved O, Rhule E (2014) Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird *Harmonia axyridis*. *Molecular ecology*, **23**, 5979–5997. <https://doi.org/10.1111/mec.12989>
- Loxdale HD, Hardie JIM, Halbert S, Footitt R, Kidd NA, Carter CI (1993) The relative importance of short- and long-range movement of flying aphids. *Biological Reviews*, **68**, 291–311. <https://doi.org/10.1111/j.1469-185X.1993.tb00998.x>
- Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
- Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M (2014) Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ*, **2**, e593. <https://doi.org/10.7717/peerj.593>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, **17**, 10–12. <https://doi.org/10.14806/ej.17.1.200>
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, **351**, 652–654. <https://doi.org/10.1038/351652a0>
- Meirmans PG, Tienderen PHV (2004) genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794. <https://doi.org/10.1111/j.1471-8286.2004.00770.x>
- Menozi P, Piazza A, Cavalli-Sforza L (1978) Synthetic maps of human gene frequencies in Europeans. *Science*, **201**, 786–792. <https://doi.org/10.1126/science.356262>
- Mizrahi-Man O, Davenport ER, Gilad Y (2013) Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. *PLOS ONE*, **8**, e53608. <https://doi.org/10.1371/journal.pone.0053608>
- Morales-Hojas R, Sun J, Iraizoz FA, Tan X, Chen J (2020) Contrasting population structure and demographic history of cereal aphids in different environmental and agricultural landscapes. *Ecology and Evolution*, **10**, 9647–9662. <https://doi.org/10.1002/ece3.6565>
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press.
- Oliver K, Degnan P, Burke G, Moran N (2009) Oliver KM, Degnan PH, Burke GR, Moran NA. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu Rev Entomol* 55: 247-266. *Annual review of entomology*, **55**, 247–66. <https://doi.org/10.1146/annurev-ento-112408-085305>
- Peccoud J, Figueroa CC, Silva AX, Ramirez CC, Mieuze L, Bonhomme J, Stoeckel S, Plantegenest M, Simon J-C (2008) Host range expansion of an introduced insect pest through multiple colonizations of specialized clones. *Molecular Ecology*, **17**, 4608–4618. <https://doi.org/10.1111/j.1365-294X.2008.03949.x>
- Piffaretti J, Clamens A-L, Vanlerberghe-masutti F, Gupta RK, Call E, Halbert S, Jouselin E (2013) Regular or covert sex defines two lineages and worldwide superclones within the leaf-curl plum aphid (*Brachycaudus helichrysi*, Kaltenbach). *Molecular Ecology*, **22**, 3916–3932. <https://doi.org/10.1111/mec.12371>
- Piry S, Luikart G, Cornuet J-M (1999) Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity*, **90**, 502–503. <https://doi.org/10.1093/jhered/90.4.502>
- Popkin M, Piffaretti J, Clamens A-L, Qiao G-X, Chen J, Vitalis R, Vanlerberghe-Masutti F, Gupta RK, Lamaari M, Langella O, Coeur d’acier A, Jouselin E (2017) Large-scale phylogeographic study of the cosmopolitan aphid pest *Brachycaudus helichrysi* reveals host plant associated lineages that evolved in allopatry. *Biological Journal of the Linnean Society*, **120**, 102–114. <https://doi.org/10.1111/bij.12869>
- Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in ecology & evolution*, **16**, 37–45. [https://doi.org/10.1016/S0169-5347\(00\)02026-7](https://doi.org/10.1016/S0169-5347(00)02026-7)
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945. <https://doi.org/10.1534/genetics.116.195164>

- Prugnolle F, Manica A, Balloux F (2005) Geography predicts neutral genetic diversity of human populations. *Current Biology*, **15**, R159–R160. <https://doi.org/10.1016/j.cub.2005.02.038>
- Pudlo P, Marin J-M, Estoup A, Cornuet J-M, Gautier M, Robert CP (2016) Reliable ABC model choice via random forests. *Bioinformatics*, **32**, 859–866. <https://doi.org/10.1093/bioinformatics/btv684>
- Puechmaille SJ (2016) The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular ecology resources*, **16**, 608–627. <https://doi.org/10.1111/1755-0998.12512>
- Raynal L, Marin J-M, Pudlo P, Ribatet M, Robert CP, Estoup A (2019) ABC random forests for Bayesian parameter inference. *Bioinformatics*, **35**, 1720–1728. <https://doi.org/10.1093/bioinformatics/bty867>
- Rebijith KB, Asokan R, Hande HR, Joshi S, Surveswaran S, Ramamurthy VV, Krishna Kumar NK (2017) Reconstructing the macroevolutionary patterns of aphids (Hemiptera: Aphididae) using nuclear and mitochondrial DNA sequences. *Biological Journal of the Linnean Society*, **121**, 796–814. <https://doi.org/10.1093/biolinnean/blx020>
- Ripley BD, Ripley BD The R project in statistical computing. *MSOR Connections*, 2001.
- Rispe C, Legeai F, Nabity PD, Fernández R, Arora AK, Baa-Puyoulet P, Banfill CR, Bao L, Barberà M, Bouallègue M, Bretaudeau A, Brisson JA, Calevro F, Capy P, Catrice O, Chertemps T, Couture C, Delière L, Douglas AE, Dufault-Thompson K, Escuer P, Feng H, Forneck A, Gabaldón T, Guigó R, Hilliou F, Hinojosa-Alvarez S, Hsiao Y, Hudaverdian S, Jacquin-Joly E, James EB, Johnston S, Joubard B, Le Goff G, Le Trionnaire G, Librado P, Liu S, Lombaert E, Lu H, Maibèche M, Makni M, Marcet-Houben M, Martínez-Torres D, Meslin C, Montagné N, Moran NA, Papura D, Parisot N, Rahbé Y, Lopes MR, Ripoll-Cladellas A, Robin S, Roques C, Roux P, Rozas J, Sánchez-Gracia A, Sánchez-Herrero JF, Santesmasses D, Scatoni I, Serre R-F, Tang M, Tian W, Umina PA, van Munster M, Vincent-Monégat C, Wemmer J, Wilson ACC, Zhang Y, Zhao C, Zhao J, Zhao S, Zhou X, Delmotte F, Tagu D (2020) The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC Biology*, **18**, 90. <https://doi.org/10.1186/s12915-020-00820-5>
- Rius M, Darling J (2014) How important is intraspecific genetic admixture to the success of colonising populations? *Trends in ecology & evolution*, **29**. <https://doi.org/10.1016/j.tree.2014.02.003>
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, **4**, e2584. <https://doi.org/10.7717/peerj.2584>
- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular ecology resources*, **8**, 103–6. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Roux C, Pannell JR (2015) Inferring the mode of origin of polyploid species from next-generation sequence data. *Molecular Ecology*, **24**, 1047–1059. <https://doi.org/10.1111/mec.13078>
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, **34**, 3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Russell JA, Weldon S, Smith AH, Kim KL, Hu Y, Łukasik P, Doll S, Anastopoulos I, Novin M, Oliver KM (2013) Uncovering symbiont-driven genetic diversity across North American pea aphids. *Molecular Ecology*, **22**, 2045–2059. <https://doi.org/10.1111/mec.12211>
- Simon J-C, Rispe C, Sunnucks P (2002) Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, **17**, 34–39. [https://doi.org/10.1016/S0169-5347\(01\)02331-X](https://doi.org/10.1016/S0169-5347(01)02331-X)
- Simon J-C, Stoeckel S, Tagu D (2010) Evolutionary and functional insights into reproductive strategies of aphids. *Comptes Rendus Biologies*, **333**, 488–496. <https://doi.org/10.1016/j.crvi.2010.03.003>
- Sow A, Brévault T, Benoit L, Chapuis M-P, Galan M, Coeur D'acier A, Delvare G, Sembène M, Haran J (2019) Deciphering host-parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding. *Scientific Reports*, **9**, 3646. <https://doi.org/10.1038/s41598-019-40243-z>
- Spengler RN (2019) Origins of the Apple: The Role of Megafaunal Mutualism in the Domestication of Malus and Rosaceous Trees. *Frontiers in Plant Science*, **10**. <https://doi.org/10.3389/fpls.2019.00617>
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stekolshchikov AV (2006) Aphids of the genus *Dysaphis* Börner (Homoptera, Aphididae) living on plants of the family polygonaceae. *Entomological Review*, **86**, 787–805. <https://doi.org/10.1134/S0013873806070049>

- Stukenbrock EH (2016) The role of hybridization in the evolution and emergence of new fungal plant pathogens. *Phytopathology*, **106**, 104–112. <https://doi.org/10.1094/PHYTO-08-15-0184-RVW>
- Stukenbrock EH, McDonald BA (2008) The Origins of Plant Pathogens in Agro-Ecosystems. *Annual Review of Phytopathology*, **46**, 75–100. <https://doi.org/10.1146/annurev.phyto.010708.154114>
- Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504. <https://doi.org/10.1093/bioinformatics/btn478>
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595. <https://doi.org/10.1093/genetics/123.3.585>
- Tardío J, Arnal A, Lázaro A (2021) Ethnobotany of the crab apple tree (*Malus sylvestris* (L.) Mill., Rosaceae) in Spain. *Genetic Resources and Crop Evolution*, **68**, 795–808. <https://doi.org/10.1007/s10722-020-01026-y>
- Tsuchida T, Koga R, Shibao H, Matsumoto T, Fukatsu T (2002) Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, *Acyrtosiphon pisum*. *Molecular Ecology*, **11**, 2123–2135. <https://doi.org/10.1046/j.1365-294X.2002.01606.x>
- Turcotte MM, Araki H, Karp DS, Poveda K, Whitehead SR (2017) The eco-evolutionary impacts of domestication and agricultural practices on wild species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**, 20160033. <https://doi.org/10.1098/rstb.2016.0033>
- Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935. <https://doi.org/10.1046/j.1365-294X.2004.02076.x>
- Vorburger C, Lancaster M, Sunnucks P (2003) Environmentally related patterns of reproductive modes in the aphid *Myzus persicae* and the predominance of two ‘superclones’ in Victoria, Australia. *Molecular Ecology*, **12**, 3493–3504. <https://doi.org/10.1046/j.1365-294X.2003.01998.x>
- Wang Y, Hereward J, Zhang G (2015) High Spatial Genetic Structure and Genetic Diversity in Chinese Populations of *Sitobion miscanthi* (Hemiptera: Aphididae). *Journal of economic entomology*, **109**. <https://doi.org/10.1093/jee/tov294>
- Warneys R, Gaucher M, Robert P, Aligon S, Anton S, Aubourg S, Barthes N, Braud F, Cournol R, Gadenne C (2018) Acibenzolar-S-Methyl Reprograms Apple Transcriptome Toward Resistance to Rosy Apple Aphid. *Frontiers in plant science*, **9**, 1795. <https://doi.org/10.3389/fpls.2018.01795>
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9)
- Watterson GA (1978) The homozygosity test of neutrality. *Genetics*, **88**, 405–417.
- Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC bioinformatics*, **11**, 1–7. <https://doi.org/10.1186/1471-2105-11-116>
- Wei G, Zuurui S, Zhihong L, Lingwang G (2005) Migration and population genetics of the grain aphid *Macrosiphum miscanti* (Takahashi) in relation to the geographic distance and gene flow. *Progress in Natural Science*, **15**, 1000–1004. <https://doi.org/10.1080/10020070512331343176>
- Weir B, Cockerham C (1984) Weir BS, Cockerham CC.. Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* 38: 1358-1370. *Evolution*, **38**, 1358–1370. <https://doi.org/10.2307/2408641>
- Wilkanić B (1993) The influence of feeding of the rosy apple aphid, *Dysaphis plantaginea* (Pass.) (Homoptera: Aphididae) on the growth of apple fruits. *Roczniki Nauk Rolniczych. Seria E, Ochrona Roślin*, **23**, 75–78.
- Xiang Y, Huang C-H, Hu Y, Wen J, Li S, Yi T, Chen H, Xiang J, Ma H (2017) Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Molecular biology and evolution*, **34**, 262–281. <https://doi.org/10.1093/molbev/msw242>
- Zepeda-Paulo F, Ortiz-Martínez S, Silva AX, Lavandero B (2018) Low bacterial community diversity in two introduced aphid pests revealed with 16S rRNA amplicon sequencing. *PeerJ*, **6**, e4725. <https://doi.org/10.7717/peerj.4725>
- Zepeda-Paulo FA, Simon J -C., Ramírez CC, Fuentes-Contreras E, Margaritopoulos JT, Wilson ACC, Sorenson CE, Briones LM, Azevedo R, Ohashi DV, Lacroix C, Glais L, Figueroa CC (2010) The invasion route for an insect pest species: the tobacco aphid in the New World. *Molecular Ecology*, **19**, 4738–4752. <https://doi.org/10.1111/j.1365-294X.2010.04857.x>

- Zhang B, Edwards O, Kang L, Fuller S (2014) A multi-genome analysis approach enables tracking of the invasion of a single Russian wheat aphid (*Diuraphis noxia*) clone throughout the New World. *Molecular Ecology*, **23**, 1940–1951. <https://doi.org/10.1111/mec.12714>
- Zhou H-X, Zhang R-M, Tan X-M, Tao Y-L, Wan F, Wu Q, Chu D (2015) Invasion Genetics of Woolly Apple Aphid (Hemiptera: Aphididae) in China. *Journal of Economic Entomology*, **108**. <https://doi.org/10.1093/jee/tov074>
- Zytynska SE, Weisser WW (2016) The natural occurrence of secondary bacterial symbionts in aphids. *Ecological Entomology*, **41**, 13–26. <https://doi.org/10.1111/een.12281>

6. OUTLINE CHAPTER 3 : GENOMIC BASES OF LOCAL ADAPTATION IN THE ROSY APPLE APHID (*DYSAPHIS PLANTAGINEA*) TO ITS HOST, THE CULTIVATED APPLE (*MALUS DOMESTICA*)

The genomic patterns of local adaptation have been analyzed by scanning for highly diverged genomic regions among populations from different environments (Andolfatto, 2001; Nielsen, 2005). This pattern of genomic divergence can be due to the spread of advantageous alleles in a given population, generating a transient reduction in neutral genetic diversity at sites close to the selected locus (selective sweep; Stephan, 2019), leading to the divergence in these regions of this population from other populations. Genomic patterns of local adaptation have been investigated in plant pathogens (Stukenbrock and Bataillon, 2012), such as in the model systems *Microbotryum lychnidis-dioicae* and *Microbotryum silenes-dioicae* (Badouin et al., 2017). Genomic footprints of local adaptation were also evaluated in crop pests, like *Plutella xylostella* L. (Chen et al., 2021), *Phaulacridium vitatum* Sjöstedt (Yadav et al., 2019), or *Grapholita molesta* (Cao et al., 2022), revealing candidate regions under positive selection. In aphids, studies on the model species *Acyrtosiphon pisum* revealed candidate genes under positive selection related to *cathepsin B-like* proteins with proteolytic activity to face the plant response (Review in Simon et al., 2015). In another study on *A. pisum*, signatures of recent positive selection are also observed in olfactory and gustatory receptors related to host specialization (Smadja et al., 2009). Genome scans in different soybean aphid biotypes (*Aphis glycines* Matsumura) revealed selective sweeps in effector genes (Coates et al., 2020). However, there is limited information on the genomic pattern of local adaptation of aphids to their environment.

In Chapter 3, I investigated the genomic bases of (local) adaptation in the rosy apple aphid. We used a representative sample (126 individuals) of *D. plantaginea* samples representing the population structure and diversity observed in Vazquez et al. 2021 (Chapter 2) and added individuals from Canada and Israel. We also included samples of *D. devectora* (Romania and Iran) and *D. pyri* (France, Spain, and Iran), two related species found on the apple and pear trees, respectively. We extracted the DNA of each aphid individual and sequenced them using the Illumina platform (2*150 bp, 30X, NovaSeq6000). SNPs were then called after mapping onto the rosy apple aphid reference genome available in the team. We aimed to answer the following questions: Do the three species, *D. plantaginea*, *D. devectora*, and *D. pyri*, form distinct genetic groups? What is the population genetic diversity and structure in *D. plantaginea*? We ask these two questions as the knowledge of the population genetic structure and diversity is essential to avoid false positive signals of candidate genes for adaptation (Li et al., 2012; Tiffin and Ross-Ibarra, 2014; Hoban et al., 2016). Eventually, is there any signature of positive selection in the genomes in each population that could reflect adaptation or local adaptation?

After several filtering steps, we used a Variant Calling File (VCF) file excluding SNPs in linkage disequilibrium to infer the aphid population genetic structure. We obtained 49,965 unlinked SNPs and found a clear genetic structure separating the three aphid species: *D. devectora*, *D. plantaginea*, and *D. pyri*. We then explored the population structure only in *D. plantaginea*. We found four genetic groups: North American (Canada and the US), Israeli, Moroccan, and Spanish, and another group including European countries. I detected a recent population expansion corresponding to the Holocene period and the occurrence of gene flow among the *D. plantaginea* genetic groups. We then scanned for positive selection (particularly selective sweeps) in the genomes of the six populations: *D. devectora*, *D. pyri*, and *D.*

plantaginea from North America, Israel, Morocco/Spain, and European countries. We found signals of selective sweeps with the μ statistic estimated with RAiSD (Alachiotis and Pavlidis, 2018) along the genomes of each population detected above. However, demographic scenarios, signals of selective sweeps using different statistics, and environmental association analysis should be performed to elucidate the genomic signals adaptation of the rosy apple aphid.

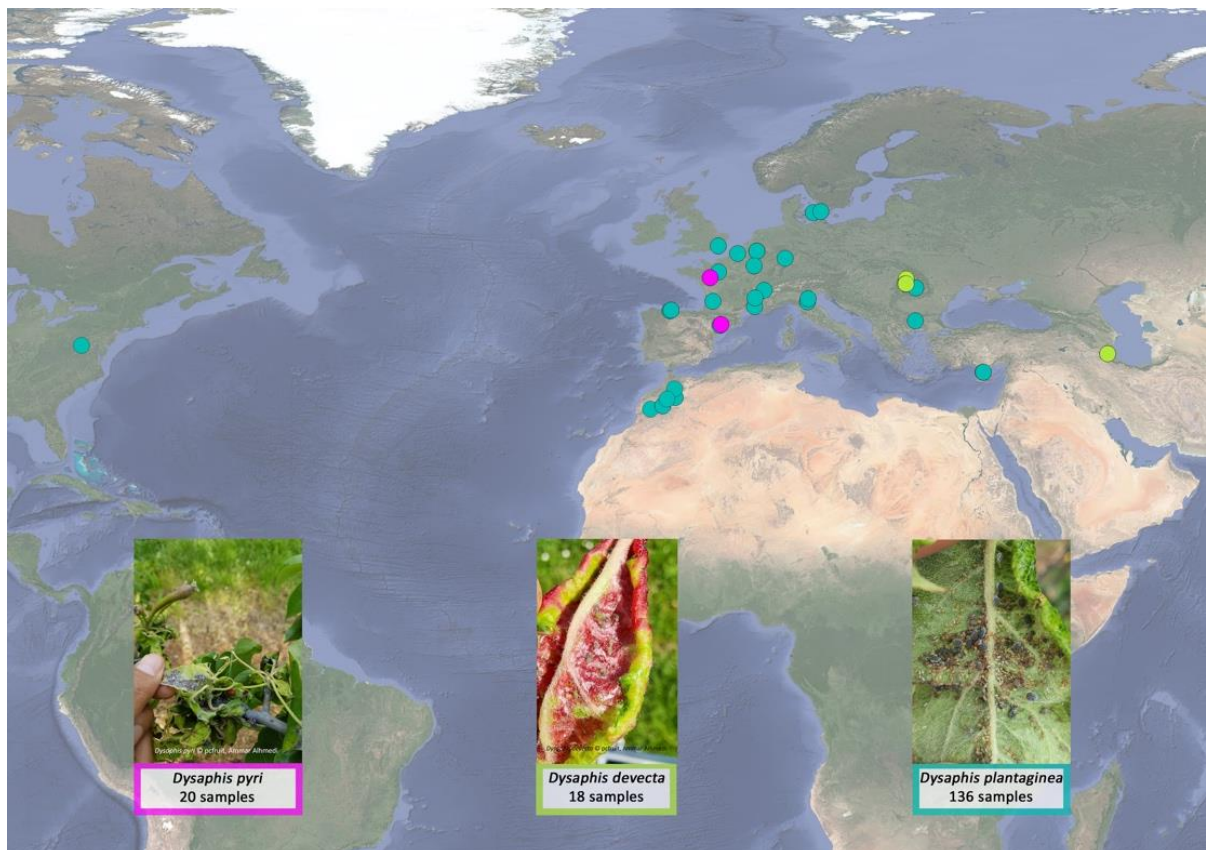


Figure 1. Samples of *Dysaphis* spp. used to scan the genomes for patterns of positive selection suggestive of (local) adaptation. We included 18 samples of *Dysaphis devectora* (green) from Romania and Iran, 136 individuals of *Dysaphis plantaginea* (blue) from Canada, Belgium, Bulgaria, Denmark, France, Germany, Italy, Israel, Morocco, Romania, Sweden, Spain, UK, and the US, and 20 Individuals of *Dysaphis pyri* (pink) from Spain, France, and Iran.

PERSPECTIVES

I will dig deeper into the rosy apple aphid demographic history compared to Chapter 2 (SSR markers study). I plan to use demography inferences using multiple sequential Markovian coalescents (MSMC2; Schiffels and Durbin, 2014).

We will confront the signatures of positive selection between the three statistics, RaisD, OmegaPlus (Alanchiotis et al., 2012), and SweeD (Pavlidis et al., 2013), and seek signatures of balancing selection in the genomes (Fijarczyk and Babik, 2015). We will also associate the SNP allele frequency with climate variables to detect SNPs associated with local adaptation to climate (Duruz et al., 2019). We will also test for the accumulation of deleterious mutations along the colonization front (Peischl et al., 2013). We are preparing these analyses for further publication in a journal.

These findings represent significant information about the genomic bases of adaptation in phytophagous insects (*i.e.*, crop pests). We also complement the results of the rosy apple aphid genomic signals of adaptation at a micro-evolutionary scale with the ecological bases of local adaptation in aphids I evaluated in Chapter 4 of this Ph.D. thesis.

OUTLINE REFERENCES

Alachiotis, N., Stamatakis, A., & Pavlidis, P. (2012). OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics*, 28(17), 2274-2275. <https://doi.org/10.1093/bioinformatics/bts419>

Alachiotis, N., & Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications biology*, 1(1), 79. <https://doi.org/10.1038/s42003-018-0085-8>

Andolfatto, P. (2001). Adaptive hitchhiking effects on genome variability. *Current opinion in genetics & development*, 11(6), 635-641. [https://doi.org/10.1016/S0959-437X\(00\)00246-X](https://doi.org/10.1016/S0959-437X(00)00246-X)

Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguileta, G., Snirc, A., ... & Giraud, T. (2017). Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Molecular Ecology*, 26(7), 2041-2062. <https://doi.org/10.1111/mec.13976>

Cao, L. J., Song, W., Chen, J. C., Fan, X. L., Hoffmann, A. A., & Wei, S. J. (2022). Population genomic signatures of the oriental fruit moth related to the Pleistocene climates. *Communications Biology*, 5(1), 142. <https://doi.org/10.1038/s42003-022-03097-2>

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1). <https://doi.org/10.1186/s13742-015-0047-8>

Chen, Y., Liu, Z., Régnière, J., Vasseur, L., Lin, J., Huang, S., ... & You, S. (2021). Large-scale genome-wide study reveals climate adaptive variability in a cosmopolitan pest. *Nature Communications*, 12(1), 7206. <https://doi.org/10.1038/s41467-021-27510-2>

Coates, B. S., Hohenstein, J. D., Giordano, R., Donthu, R. K., Michel, A. P., Hodgson, E. W., & O'Neal, M. E. (2020). Genome scan detection of selective sweeps among biotypes of the soybean aphid, *Aphis glycines*, with differing virulence to resistance to *A. glycines* (Rag) traits in soybean, *Glycine max*. *Insect Biochemistry and Molecular Biology*, 124, 103364. <https://doi.org/10.1016/j.ibmb.2020.103364>

Evans, R. C., & Campbell, C. S. (2002). The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. *American journal of botany*, 89(9), 1478-1484. <https://doi.org/10.3732/ajb.89.9.1478>

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., ... & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4), 379-397. <https://doi.org/10.1086/688018>

Huson, D. H., & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23(2), 254-267. <https://doi.org/10.1093/molbev/msj030>

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, 15(5), 1179-1191. <https://doi.org/10.1111/1755-0998.12387>

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987-2993. <https://doi.org/10.1093/bioinformatics/btr509>

Li, J., Li, H., Jakobsson, M., Li, S. E. N., Sjödin, P. E. R., & Lascoux, M. (2012). Joint analysis of demography and selection in population genetics: where do we stand and where could we go?. *Molecular ecology*, 21(1), 28-44. <https://doi.org/10.1111/j.1365-294X.2011.05308.x>

Li, J., Zhang, M., Li, X., Khan, A., Kumar, S., Allan, A. C., ... & Wu, J. (2022). Pear genetics: Recent advances, new prospects, and a roadmap for the future. *Horticulture Research*, 9, uhab040. <https://doi.org/10.1093/hr/uhab040>

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Molecular ecology resources*, 21(2), 584-595. <https://doi.org/10.1111/1755-0998.13265>

Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39, 197-218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>

Olvera-Vazquez, S. G., Remoué, C., Venon, A., Rousselet, A., Grandcolas, O., Azrine, M., ... & Cornille, A. (2021). Large-scale geography survey provides insights into the

colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa. *Peer Community Journal*, 1. <https://doi.org/10.24072/pcjournal.26>

Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular biology and evolution*, 30(9), 2224-2234. <https://doi.org/10.1093/molbev/mst112>

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573-589. <https://doi.org/10.1534/genetics.114.164350>

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature genetics*, 46(8), 919-925. <https://doi.org/10.1038/ng.3015>

Simon, J. C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquier, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in functional genomics*, 14(6), 413-423. <https://doi.org/10.1093/bfpg/elv015>

Stephan, W. (2019). Selective sweeps. *Genetics*, 211(1), 5-13. <https://doi.org/10.1534/genetics.118.301319>

Stukenbrock, E. H., & Bataillon, T. (2012). A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. <https://doi.org/10.1371/journal.ppat.1002893>

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12), 673-680. <https://doi.org/10.1016/j.tree.2014.10.004>

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... & Banks, E. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics*, 43(1), 11-10. <https://doi.org/10.1002/0471250953.bi11110s43>

Yadav, S., Stow, A. J., & Dudaniec, R. Y. (2019). Detection of environmental and morphological adaptation despite high landscape genetic connectivity in a pest grasshopper (*Phaulacridium vittatum*). *Molecular Ecology*, 28(14), 3395-3412 <https://doi.org/10.1111/mec.15146>

6.1 GENOMIC BASES OF LOCAL ADAPTATION IN THE ROSY APPLE APHID (*DYSAPHIS PLANTAGINEA*) TO ITS HOST, THE CULTIVATED APPLE (*MALUS DOMESTICA*) (MANUSCRIPT IN PREPARATION)

Olvera-Vazquez Sergio Gabriel, Xilong Chen, Christelle Buchard, Nathalie Rodde, Claire Mottet, Alexandre Degrave, Elorri Segura, Lior Gur, Moshe Reuveni, Ammar Ahlmedi, AnaMaria Roman, Vincent Arnaud, Michel Dron, Kim H., Lauren W., Elisa Marchetti, Stefano Tartarini, Neus Rodríguez-Gasol, Jean Engelman, Jabraeil Razmjou, Karsten Mody, Ludovic Tournant, Marcos Minarro Prado, Morgane Roth, Pierre Franck, Radoslav Andreev, Stine Kramer Jacobsen, Youssef Khachtib, Georgina Alins, William Marande, Fabrice Legeai, Amandine Cornille

ABSTRACT

Assessing the population genetic diversity, structure, and demographic history is critical to understanding the evolutionary processes underlying rapid colonization and range shifts of phytophagous insects. Detecting genomic regions related to the adaptation to the abiotic (*i.e.*, light, temperature) and/or biotic (*i.e.*, host) environment of phytophagous insects is crucial to mitigate current and future pest outbreaks. Candidate genes related to biotic or abiotic environments through different genomic patterns (*e.g.*, hard sweep, soft sweep, balancing selection) have been reported in phytophagous insect genomes. However, limited studies are available for aphids, one of the phytophagous insects that contain important crop pests. Moreover, more information is needed on the genes involved in the adaptation of the *Dysaphis* genus to their hosts and environment. *Dysaphis* genus includes *Dysaphis devectora*, *Dysaphis plantaginea*, and *Dysaphis pyri*, important pests of the cultivated apple and pear. Here, we investigated genomic signatures of adaptation using population genomics on short-read sequenced samples of the three species for the first time. We detected a clear split among *D. devectora*, *D. plantaginea*, and *D. pyri* and four main panmictic groups within *D. plantaginea* (American, Moroccan/Spanish, Israeli, and European). We detected a recent population expansion of the *D. plantaginea* genetic groups with gene flow

between the European and Moroccan/Spanish group and the American and European group. We present preliminary results of genome scans of the six genetic groups, which revealed candidate outlier SNPs, but further investigations are needed. We will also seek SNPs associated with the climate variables. Besides, additional information about the demography history are need to confirm colonization scenarios. The findings of this study will bring insight into the role of selection and demography in shaping the divergence of aphid populations and species in response to their abiotic and biotic environments. This information will help to develop strategies for sustainable pest management and understand the migration of crop pests.

Keywords: local adaptation, aphid, genome scans, positive selection, gene flow, pest, divergence.

6.2 INTRODUCTION

Elucidating the effect of demography and selection in modeling population genetic variation allows us to understand the evolutionary processes involved in divergence (Via, 2009). Assessing population genetic diversity, structure, and demographic history is also critical to understanding the evolutionary processes underlying rapid colonization and range shifts (Estoup & Guillemaud, 2010; Lawson Handley et al., 2011; Li et al., 2012; Tiffin and Ross-Ibarra, 2014; Hoban et al., 2016; Fraimout et al., 2017; Turcotte et al., 2017). The knowledge of population structure is also preliminary information needed to avoid the detection of false positives when inferring the role of selection in shaping population divergence Li et al., 2012; Tiffin and Ross-Ibarra, 2014; Hoban et al., 2016). The role of selection in the adaptation of populations to their environment can then be assessed by detecting genomic regions under selection related to their abiotic (*i.e.*, light, temperature)

and/or biotic (*i.e.*, host) environments (Li et al., 2012; Tiffin and Ross-Ibarra, 2014; Hoban et al., 2016; Nooten and Hughes, 2017).

Positive selection, resulting in a pattern of hard and soft sweeps and balancing selection, can shape the architecture of parasite genomes (Ebert and Fields, 2020; Märkle and Tellier, 2020), including phytophagous insects (Simon et al., 2015). Episodes of a quick spread of an advantageous allele in a population, called a selective sweep, can generate a transient reduction in genetic diversity at loci close to the locus under selection (Cutter and Payseur, 2013; Stephan, 2019). The quick fixation of a new beneficial mutation from a single haplotype in a population is called a hard sweep (Burke, 2012). A soft sweep occurs when multiple copies of a beneficial mutation become established and fixed together in a population (Messer and Petrov, 2013; Hahn, 2019). The main difference between hard and soft selective sweep is the expected number of different haplotypes presenting the beneficial mutation (Hermisson and Pennings, 2017; Jensen, 2014). Balancing selection implies that more than one allele is kept at a locus due to the fitness of the heterozygotes being higher than the homozygote at a locus (called overdominance) or from frequency-dependent selection with the advantage of rare alleles (Charlesworth, 2006; Siewert and Voight, 2017). Scanning the genomes for these footprints is a way to understand their relative contribution of positive and balancing selection to parasite adaptation. Note that in this preliminary draft, I will only focus on the selective sweep, but I plan to investigate soft and balancing selection patterns, too.

Phytophagous insects are ideal study subjects in the rich theoretical framework for local adaptation in host–parasite interactions (Gloss et al., 2016; Greischar & Koskella, 2007). Genomic patterns of local adaptation have been investigated in phytophagous insects (Song et al., 2015; Crossley et al., 2017; Fiteni et al.,

2022). In the lepidopteran crop pests such as diamondback moth or *Plutella xylostella* L., 3,648 SNPs associated with climatic variables are candidate loci for local adaptation to climate (Chen et al., 2021). Genome scans in the lepidopteran pest, the oriental moth or *Grapholita molesta* Busck, detected mutations in the circadian clock gene (*Clk*) and target-sites pesticide candidate genes comparing populations from Sichuan with two other populations in Yunnan in China (Cao et al., 2022). In the grasshopper crop pest *Phaulacridium vittatum* Sjöstedt, genomic patterns of local adaptation were found with 242 unique SNPs associated with latitude, mean annual temperature, and body size (Yadav et al., 2019). In the aphid crop pest *Acyrtosiphon pisum* Harris, candidate genes under selection in different biotypes were detected, some related to olfactory receptors (65% outliers SNPs) and gustatory receptors (30% outliers SNPs; Smadja et al., 2012). In another study on soybean aphid, *Aphis glycines* Matsumura, selective sweeps were detected in effector genes in different aphid biotypes and related to the interaction between these phytophagous insects and their host, the soybean (Coates et al., 2020). Therefore, these studies indicate footprints of positive selection in the genomes of phytophagous insects associated with host or abiotic environments. In addition, the population genetic structure of phytophagous insects can be due to the host specificity (Berlocher and Feder, 2002; Drès and Mallet, 2002), spatial variation, or abiotic parameters (McPherson et al., 1988; Manel et al., 2009; MacDonald et al., 2022), suggesting the role of the local host and abiotic environment in phytophagous insect evolution. However, studies on genomic footprints of local adaptation across a large geographic distribution are needed to understand better genes involved in the local adaptation of phytophagous insects to their host and/or environmental factors. To disentangle the role of selection in the evolution of phytophagous insects, the knowledge of its population structure, diversity, and de-

mographic history (divergence time, migration rates among populations, and effective population size) is also essential (Li et al., 2012; Tiffin and Ross-Ibarra, 2014; Hoban et al., 2016) to avoid false positives while testing for local adaptation.

The genus *Dysaphis* Börner contains species that are important pests of the subfamily Maloideae C. Weber (Blackman and Eastop, 2000), which includes important agricultural fruit trees, such as the cultivated apple and pear (Phipps et al., 1990; Evans and Campbell, 2002). The rosy apple aphid or *Dysaphis plantaginea* Passerini represents the major pest of the cultivated apple, while the pear-bedstraw or *Dysaphis pyri* Boyer de Fonscolombe is an important pest of the cultivated pear (Kehrli and Wyss, 2001; Evans et al., 2008). These species present host alternate (heteroecious life cycle) between a woody primary host and an herbaceous secondary host, giving one sexual generation and several asexual generations per year (holocyclic cycle; Blackman and Eastop, 2000). For *D. plantaginea*, the cultivated apple is the primary host, while *Plantago* L. is the secondary host (Bloomers et al., 2004). In the case of *D. pyri*, the cultivated pear is the primary host, while the *Galium* L. species represent the secondary host (Atlihan et al., 2017). On the other hand, the rosy leaf-curling apple aphid or *Dysaphis devecta* Walker represents a minor pest to the apple tree and presents a monoecious cycle staying in the apple tree (Gratwick, 1992), infesting older apple trees, producing characteristics of red pseudo galls by rolling the apple leaf laterally (apex to the leaf base), usually staying in the same tree after a year (Forrest and Dixon, 1975). Previous population genetics studies used microsatellite markers and revealed the population structure and demographic history of the rosy apple aphid in Europe, the Maghreb, and the US (Guillemaud et al., 2011; Olvera-Vazquez et al., 2021). The rosy apple aphid colonized quickly and recently cultivated host in Europe, the Maghreb, and the US, exchanging substantial gene flow among populations. However, samples from the Middle East, a supposed origin of the rosy apple aphid, were lacking, as

well as related *Dysaphis* species found on other hosts to track the evolutionary travel of this aphid species. The rosy apple aphid may result from a host shift from another host, the cultivated pear, on which a sister species, *Dysaphis pyri*, is currently found. The scan of genomes for islands of differentiation among *Dysaphis* species could help us to understand the genes involved in speciation in aphids and the origin of *D. plantaginea*. In addition, we have observed a variation in hatching across Europe in *D. plantaginea*, with later and earlier hatching in Northern and Southern Europe, respectively. Whether this phenotypic variation reflects a local adaptation of *D. plantaginea* populations to the abiotic or biotic conditions following its recent colonization of Europe remains unclear. Moreover, the cultivated apple displays considerable variation in sensitivity and resistance to *D. plantaginea* attacks (Miñarro et al., 2005; Pagliarani et al., 2016), and *D. plantaginea* is very rarely seen on its wild European relative, *Malus sylvestris*. Then, there is a need to evaluate the genomic bases behind the phenological variation and potentially local adaptation of the rosy apple aphid to its host, the apple tree, or its local environment.

Here, we investigated the genomic bases of (local) adaptation in a major aphid pest of the cultivated apple using a comprehensive sampling of the species and populations across their distribution and sequenced using short-read Illumina. We aimed to answer the following questions: Do the three species form distinct genetic groups? What is the population genetic diversity and structure in *D. plantaginea*? We described the population structure and diversity as it is essential to avoid detecting false positive candidate genes for adaptation (Li et al., 2012; Tiffin and Ross-Ibarra, 2014; Hoban et al., 2016). Eventually, is there any signature of positive selection in the genomes of the different aphid populations that could reflect adaptation or local adaptation?

6.3 MATERIAL AND METHODS

SAMPLING, AMPLIFICATION, AND SEQUENCING

The 174 aphid colonies used in this study were sampled in North America, Europe, and the Middle East and on two hosts (Table S1). We sampled 20 colonies of the pear-bedstraw aphid (*D. pyri*) on cultivated pear trees (*Pyrus communis* L) from Spain, France, and Iran, 18 colonies of the rosy leaf-curling apple aphid (*D. devecta*) from Romanian and 136 colonies of the rosy apple aphid (*D. plantagine*) colonies on cultivated apples (*Malus domestica* Borkh) trees in North America (Canada and USA), Europe (Belgium, Bulgaria, Denmark, France, Germany, Italy, Romania, Spain, Sweden, and the UK), North Africa (Morocco), and Iran (see details in Table S1). We sampled only one colony per tree to ensure that the sample did not contain different clones, which can occur on the same tree. Each colony was kept in ethanol (96%) at -20°C until DNA extraction. We isolated the DNA from a single apterous adult individual per colony using a previously standardized protocol (Olvera-Vazquez et al., 2021), with the addition of putting our samples under -80°C one night before the DNA extraction. We sent the DNA samples to Novogene© for whole genome paired-end resequencing (2*150 bp, 30X) using the Illumina platform (NovaSeq6000).

SINGLE NUCLEOTIDE POLYMORPHISM (SNP) MAPPING AND CALLING

We checked the quality of the raw reads using FastQC v0.11.17 (Andrews, 2010). We trimmed the sequences to bp and removed the sequences with a length of less than 70 bp using the fastp pre-processing tool v0.19.4 (Chen et al., 2018). We aligned the individual reads to the reference genome v4 of the new genome chromosome-level assembly of *D. plantaginea* (<https://bipaa.genouest.org/is/>; details about the assembly using the bwa-mem2 -t 4. We removed the redundant reads (*i.e.*, PCR or optical duplicates) from the Binary Alignment Maps (BAM) file

with MarkDuplicates from the Genome Analysis Toolkit (GATK v41.7.0; Van der Auwera et al., 2013) and SAMtools v1.9 (Li, 2011). We performed the SNP calling using HaplotypeCaller from the GATK pipeline, setting the ploidy on two (Supplementary text S1 Figure 1). We combined the outputs and obtained the raw variant call format (VCF) file for the aphid dataset with CombineGVCFs and GenotypeGVCFs from the GATK pipeline. We applied this pipeline separately to call the raw SNP VCF file for the obligate bacteria *Buchnera aphidicola* (121,311 SNPs and 174 individuals) and mitochondrial (2,654 SNPs and 174 individuals) genomes and obtained two VCF files. For the aphid genome VCF file, we filtered SNPs corresponding to the bacteria and the mitochondria genomes, and also with high missing rates, low-quality variants (*i.e.*, sites with missing information), and indels (insertion-deletions) with BCFtools v1.14 (Danecek *et al.*, 2021) and the VariantFiltration and SelectVariants modules of the GATK pipeline. We defined the cut-off for missing SNPs by plotting the distribution of missing rate per sample using the R packages ggplot2 (Wickham, 2016), dplyr (Wickham *et al.*, 2022), and cowplot (Wilke, 2020) in the R software v. 4.1.2 (R Core Team, 2021). We used a missing rate cut-off of 10% and removed samples with a depth of coverage below five or above 100 using BCFtools to minimize the occurrence of unreliable genotypes. We did further SNP filtering for population structure and genetic diversity analyses to keep only SNPs reflecting population history to avoid bias associated with confounding effects between demography and selection. We removed non-informative SNPs, with > 20% of missing data or with minor allele frequency (MAF) < 0.05 using BCFtools. We used the pairwise KING-robust kinship cut-off to detect duplicate or closely genetically related individuals, setting a value of 0.354 with BCFtools. We also filtered non-synonymous and linked SNPs. To that aim, we first annotated SNPs with SnpEff v5.1 (Cingolani *et al.*, 2012) and then filtered the non-synonymous SNPs using BCFtools v1.9 and SAMtools v1.9 (Danecek *et al.*, 2021). We also removed the remaining SNPs in linkage disequilibrium (*LD*). We estimated

the *LD* decay with PopLDdecay (Zhang *et al.*, 2018) using the squared correlation coefficient (r^2) between pairs of SNPs. We evaluated the decayed physical distance between SNPs as the distance at which the maximum r^2 is halved. We plotted the decay of r^2 with R (Figure S1). We, therefore, set an average *LD* decay of 1 Kb to filter the potentially linked SNPs and retain SNPs separated by a higher distance than the estimated *LD* decay.

POPULATION STRUCTURE INFERENCES, GENETIC DIFFERENTIATION, AND DIVERSITY

We inferred the aphid population genetic structure using a Bayesian framework to infer panmictic groups for a predefined number of genetic clusters (K) implemented in the fastSTRUCTURE software v1.0 (Raj *et al.*, 2014). We ran fastSTRUCTURE using the logarithm function, including the three species ($N=143$) and only *D. plantaginea*, as we had more samples for this species ($N=118$). We set the number of populations from $K=2$ to $K=20$ with 20 repetitions per K . We generated a consensus solution for each K value using CLUMPAK (Kopelman *et al.*, 2015). We used StructureSelector (<http://lmme.qdio.ac.cn/StructureSelector/>; Li and Li, 2017) to select the optimal K value and use the *Puechemaille* threshold value 0.5. We plot the spatial genetic structure with QGIS 3.12 “Las Palmas” (<http://qgis.org/>) using the membership coefficients inferred with fastSTRUCTURE averaged per sampled sites. The bar plots were obtained with PopHelper Shiny (Francis, 2017).

We evaluated the genetic differentiation among populations detected by fastSTRUCTURE using three analyses. First, we used a principal component analysis (PCA), generating the eigenvalues using the Plink v2 (Chang *et al.*, 2015). Then, we plot the first three principal components with the R package ggplot2 (Wickham, 2009). Individuals assigned to a given cluster with a membership coefficient ≥ 0.90

were color-coded according to the respective color of each group detected with fastSTRUCTURE. We chose this threshold based on the distribution of the maximum membership coefficients inferred with fastSTRUCTURE (see Results). We then analyzed the genetic relationships among individuals using two approaches. We estimated the Nei's distance between individuals to compute the neighbor-net tree using SplitsTree v4 (Huson and Bryant, 2005). We used IQ-TREE software (Minh et al., 2020) to infer the genetic relationships using a maximum likelihood tree model finder option, ultrafast bootstrap 1,000, and 1,000 branch support tests (SH-aLRT).

We computed descriptive population genetic estimates for each population (*i.e.*, each cluster inferred with fastSTRUCTURE, excluding admixed individuals with a membership coefficient <0.90 to a given cluster; see results for details). We utilized Stacks v2.61 (Catchen *et al.*, 2013) to estimate genetic differentiation among populations with F_{ST} (genetic differentiation between populations), as well as genetic diversity indices for each population, including π (Nei's diversity, or nucleotide diversity within the population), F_{IS} (inbreeding coefficient), H_o (observed heterozygosity) and H_E (expected heterozygosity).

DEMOGRAPHIC ANALYSIS

We obtained the effective population size (N_e) and splitting time of the *D. plantaginea*, *D. pyri*, and *D. delecta* populations using the sequentially Markovian coalescent software SMC++ that incorporates estimates of recombination and LD in a site frequency spectrum framework (Terhorst et al., 2017). We utilized the reported mutation rate of *A. pisum* 2.7×10^{-10} substitutions per site (Mathers et al., 2023) and one generation time per year.

SPECIES TREE ESTIMATION AND GENE FLOW AMONG THE APHID POPULATIONS

We used SVDQuartets (Chifman and Kubatko, 2014) implemented in PAUP v.4* (Swofford, 2002) to evaluate support for all possible quartets from the datasets (49,965 unlinked SNPs), and uncertainty in relationships was quantified using 1,000 bootstrapping replicates. We also utilize this approach to gain insight into the gene flow among the aphid genetic groups that we further explore with the D-suite.

We estimated Patterson's D and f-4 ratio with the software D-suite (Malinsky et al., 2021) to infer patterns of historical gene flow between lineages. This method was based on the ABBA-BBA test, which utilizes a tree with (((P1, P2), P3), P4) shapes, the relative proportions of derived traits (*B*) shared by two lineages are compared. A form of BBAA represents a derived trait shared due to common ancestry (*i.e.*, synapomorphy), BABA incomplete lineage sorting, and ABBA gene flow between P1-P3 or P2-P3. Incomplete lineage sorting ABBAs and BABAs should occur in the same proportion across the genome while increasing ABBAs indicates gene flow between the P2 and P3 lineages. We tested two hypotheses, first gene flow among *D. plantaginea* populations (America N = 7, Israel N = 6, Europe N = 21, and Morocco and Spain N = 14) using the Israel group as an outgroup based on the genetic differentiation. The second hypothesis tested gene flow between *D. plantaginea* populations and *D. pyri* (N = 11) as an outgroup.

DETECTION OF SIGNATURES OF POSITIVE SELECTION

We seek genomic signatures of recent positive selection (*i.e.*, selective sweep) in the populations detected above (*i.e.*, cluster inferred with fastSTRUCTURE, excluding admixed individuals with a membership coefficient <0.90; see results for details). We define a selective sweep as a signal of the recent signature

of positive selection; the signal is represented by the rise in frequency of a beneficial allele in a population, resulting in lower allelic diversity and frequencies with one abundant allele (the selected variant) and an increase of LD around the selected locus. To avoid skew allelic frequencies and biased selection analyses, we removed admixed individuals (*i.e.*, individuals with a membership coefficient <0.90 to a given cluster detected with fastSTRUCTURE) with BCFtools v1.9 (Danecek *et al.*, 2021). We used a complete set of SNPs for selection analyses (7,622,598 SNPs, Supplementary text S1). We used the composite measure μ statistics implemented in RAiSD based on the expected effects of selective sweeps on nucleotide diversity, LD, and the site frequency spectrum (Alachiotis and Pavlidis, 2018). We also used Omega Plus v2.3 software to get the ω statistics that measure the changes in LD for a defined genomic region from an empirical expectation estimated from the magnitude of the genome-wide LD (Alachiotis *et al.*, 2012). Plots were generated with the ggplot2 (Wickham, 2016), cowplot (Wilke, 2020), and dplyr (Wickham *et al.*, 2022) R packages. The perspective section below describes the future work of positive selection and demography history analyses.

6.4 RESULTS

SNP DATASET

We had 174 individuals initially; after applying VCF filtering (Supplementary text 1), we obtained 7,622,598 SNPs for 143 individuals, including 14 *D. devectora*, 11 *D. pyri*, and 118 *D. plantaginea* (Figure S2). After filtering for linked and non-synonymous SNPs, we obtained 49,965 unlinked synonymous SNPs for 143 individuals to infer the aphid population genetic structure and relationships.

INTERFERENCES OF THE APHID POPULATION GENETIC STRUCTURE

fastSTRUCTURE revealed a clear population genetic structure among the three species *D. plantaginea*, *D. pyri*, and *D. devectora* samples (Figures 1A and S3) at $K=3$. The structure Selector further confirmed the optimal K value (Figure S4). Individuals were then assigned to their respective population (*i.e.*, individuals with a membership coefficient >0.90 to a given cluster detected with fastSTRUCTURE, Figure S5). Once individuals were assigned to their respective populations, we further confirmed the three main clusters in the neighbour net, phylogenetic tree, and PCA (Figures 1B, C, and Figure S6). The mitochondrial and bacterial SNPs dataset also indicated that the three species formed distinct genetic groups (Figures S7 and S8).

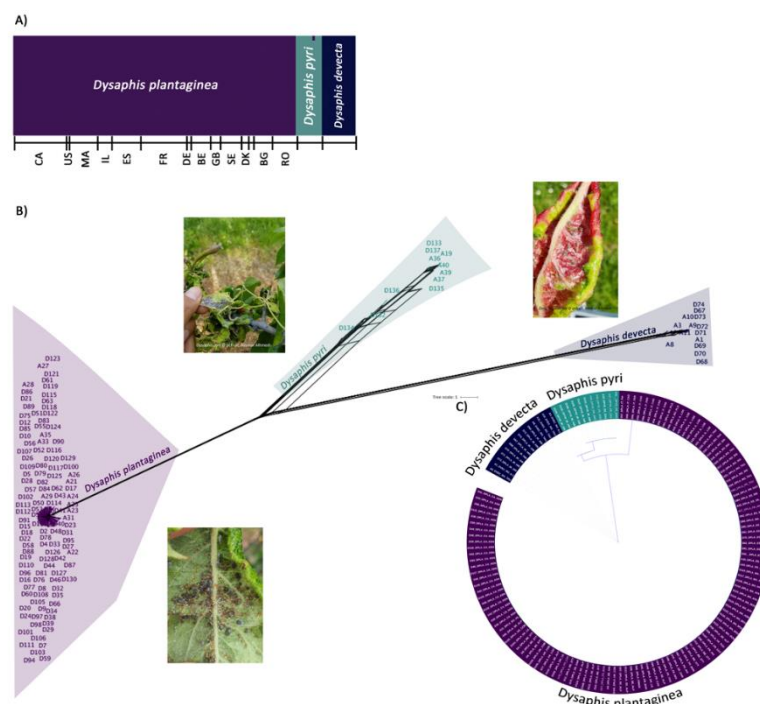


Figure 1. Population structure of *Dysaphis plantaginea*, *Dysaphis pyri*, and *Dysaphis devectora* ($N=143$ individuals and 49,965 unlinked synonymous SNPs). A) *D. plantaginea* (purple), *D. pyri* (aqua), and *D. devectora* (dark blue) detected with fastSTRUCTURE (Raj et al., 2014) at $K = 3$ based on Structure-Selector (Li and Li, 2017). B) Neighbour net among individuals built with SplitsTree (Huson and Bryant, 2005). C) Genetic relationships represented with a circular tree using IQ-TREE (Minh et al., 2020). Individuals were colored according to the respective color of each cluster detected with fastSTRUCTURE.

In *D. plantaginea* only ($N=118$), fastSTRUCTURE revealed four main genetic clusters spread in North America (teal), Morocco and Spain (yellow), Western and Eastern Europe (cyan), and Israel (light green; Figure 2A and Figures S9). The structure selector indicated an optimal $K = 9$ (Figure S10), at which four main genetic groups were observed. Individuals were then assigned to their respective population (*i.e.*, individuals with a membership coefficient >0.90 to a given cluster detected with fastSTRUCTURE at $K=9$, Figure S5). The neighbor net analysis and phylogenetic tree (Figures 2B and C) further validated the population genetic structure in *D. plantaginea*, as well as the genetic relationships among individuals built with mitochondria (Figure S11) and bacterial SNPs (Figure S12). Therefore, we assumed four main populations in *D. plantaginea* distributed in North America, Israel, Morocco and Spain, and Western and Central Europe. We further investigated the substructure within the European populations (Figure 2D), excluding the American and Israeli groups (Figure 2E). The PCA further confirmed a clustering of Moroccan and some Spanish individuals while other Europeans clustered homogeneously. A stubble clustering of Bulgarian, Romanian, and Italian individuals from the rest of the countries was observed, with a low percentage of variation explained by the axes of the PCA. The weak genetic structure in the European countries was further confirmed with an additional fastSTRUCTURE analysis, including European samples (Figures S13 and S14). The observed patterns suggest a cline in Europe, as suggested in Olvera-Vázquez et al. 2021.

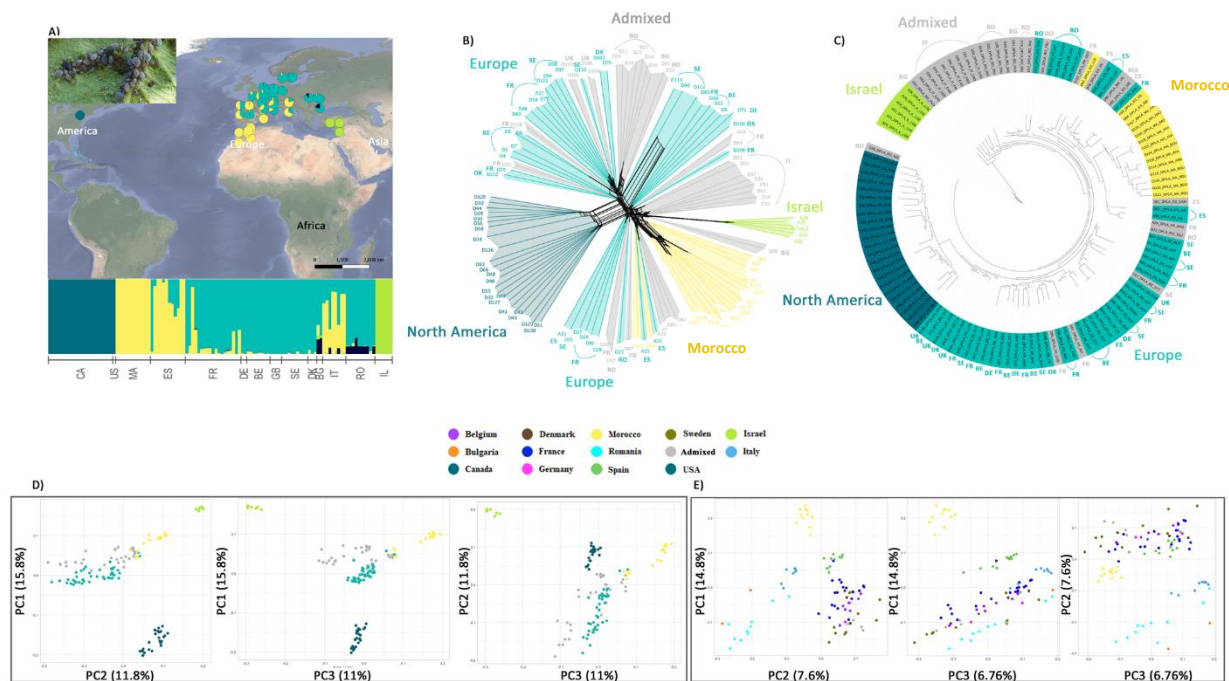


Figure 2. *Dysaphis plantaginea* population structure in the Middle East, North America, and Europe.

A) Aphid genetic structure inferred with FastSTRUCTURE at $K = 9$ ($N=118$ individuals, 49,965 unlinked synonymous SNPs). B) Neighbour net among individuals built with SplitsTree. Individuals assigned to a given cluster with a membership coefficient ≥ 0.90 were color-coded according to the respective color of each cluster, and admixed individuals (*i.e.*, individuals with a membership coefficient to any given cluster < 0.90) were colored in grey. In teal color: North America; yellow: Morocco, Spain, and Italian individuals; light green: Israel; and cyan: rest of the European countries. Genetic relationships are represented by a circular tree built with IQ-Tree (Minh et al., 2020). Individuals were colored as for the SplitsTree. D) Principal component analysis (PCA) representing the genetic variation among *D. plantaginea* individuals. The same color is used for the SplitsTree. E) Principal component analysis (PCA), representing the genetic variation excluding the Israeli *D. plantaginea* individuals. The code of the countries is the following: CA: Canada; BE: Belgium; BG: Bulgaria; DE: Germany; DK: Denmark; ES: Spain; FR: France; GB: UK; IL: Israel; IT: Italy; RO: Romania; SE: Sweden; US: The US.

GENETIC DIVERSITY IN THE *DYSAPHIS* POPULATION

We detected the lowest genetic diversity in *D. devectora* and the highest private alleles (Table 1), which may suggest that *D. devectora* is the most ancient species of the three species. *Dysaphis pyri* showed the highest genetic diversity, suggesting it may be a potential source of colonization of *D. plantaginea*. Within *D. plantaginea*, the Israeli group showed the lowest genetic diversity estimates and number of private alleles (Table 1), followed by Moroccan/Spanish and American populations. In contrast, the highest genetic diversity and number of private alleles were observed in the European population (Table 1). This may suggest waves of colonization of *D. plantaginea* out of Europe. We also found low levels of inbreeding (Mean $F_{is} = -0.015$), suggesting an excess of heterozygotes and potentially a clonal population structure.

We showed that *D. devectora* was the more genetically differentiated species, followed by *D. pyri*. Within *D. plantaginea*, we observed a low genetic differentiation of the European populations against the other three populations (North America, Israel, Morocco, and Spain; Table 1). The Israeli population was the most genetically differentiated among the *D. plantaginea* groups (Table 1).

RECENT POPULATION EXPANSION AND SPLIT OF *D. PLANTAGINEA* GENETIC GROUPS

We obtained a recent N_e of the four *D. plantaginea* (Figure 15) and *D. pyri* and *D. devectora* genetic groups (Figure 16). We observed population expansion about ten thousand years ago for the *D. plantaginea* groups from America, Israel, Morocco and Spain and *D. pyri* groups with subsequent bottlenecks.

Table 1. Genetic diversity estimates of *Dysaphis plantaginea* populations detected with fastSTRUCTURE ($N = 103$, admixed individuals were removed; 49,965 unlinked synonymous SNPs). The calculations were computed with Stacks v2.61 (Catchen *et al.*, 2013).

Aphid population	N	Sites	VS	A_p	H_o	H_e	π	F_{IS}	F_{ST}	DVEC	DPYR	Israel	Europe	Morocco and Spain	America
<i>Dysaphis devectora</i>	14	46,870	46,870	20,091	0.014	0.017	0.018	0.015	DVEC	-	0.864	0.932	0.884	0.904	0.905
<i>Dysaphis pyri</i>	11	47,722	47,722	11,253	0.035	0.054	0.057	0.063	DPYR	-	-	0.786	0.753	0.772	0.776
Israel	6	49,947	49,947	2	0.021	0.016	0.017	0.007	Israel	-	-	-	0.038	0.079	0.090
Europe	42	49,965	49,965	72	0.044	0.033	0.033	0.026	Europe	-	-	-	-	0.025	0.029
Morocco and Spain	19	49,965	49,965	1	0.029	0.024	0.025	0.008	Morocco and Spain	-	-	-	-	-	0.058
America	23	49,965	49,965	26	0.036	0.028	0.029	0.017	America	-	-	-	-	-	-

N : number of individuals assigned to the population with a membership coefficient > 0.9 ; VS: Variant sites; A_p : private alleles; H_o : observed heterozygosity; H_e : expected heterozygosity; π : nucleotide diversity within the population; F_{IS} : inbreeding coefficient; %P: percentage of polymorphism; F_{ST} : differentiation coefficient. DVEC: *Dysaphis devectora*; DPYR: *Dysaphis pyri*; Israel, Europe, Morocco and Spain, and America correspond to *Dysaphis plantaginea* groups.

The exception was *the D. devectora* group, which presented a bottleneck followed by a population expansion event with a second drastic bottleneck. We also detected the *D. plantaginea* group from Europe displayed a higher population expansion in 10^5 with a subsequent bottleneck event in short periods.

We also explored the splitting time among the aphid genetic groups. We found issues plotting the correct split time among the *D. plantaginea* genetic groups. We found the lowest event split time was 29 generations in the group from America and Europe and the highest was 113 generations from Europe and Israel. On the contrary, we found the recent split time of *D. pyri* less than 10^6 years ago and *D. devectora* more than 10^6 years ago (Figure S17).

GEN FLOW AMONG THE *D. PLANTAGINEA* GENETIC GROUPS

We obtained high bootstrap support in the aphid species tree with the SVDQuartets (Figure S18). The exception was the clade of *D. plantaginea* from America, Morocco and Spain, and Israel, presenting a low value. We detected gene flow among *D. plantaginea* groups corresponding to our hypothesis 1 (Figure S19 and Table S2). We saw gene flow between *D. plantaginea* genetic groups from Europe and Morocco and Spain, and Morocco and Spain with the American group. We also identified gene flow between *D. plantaginea* and *D. pyri* groups related to our hypothesis 2 (Figure S19 and Table S3). We found gene flow between *D. plantaginea* from America and *D. pyri* and *D. plantaginea* from Morocco and Spain and *D. pyri*.

GENOMIC SIGNATURES OF POSITIVE SELECTION OF THE SIX APHID POPULATIONS

We detected some candidate outliers suggesting positive selection using RAiSD exploring the six aphid populations (Figure 3). We observed higher signals in the six chromosomes among the different aphid populations. We also launched OmegaPlus software for each of the six populations to detect SNP candidate outliers; however, we are still defining the threshold to plot the results.

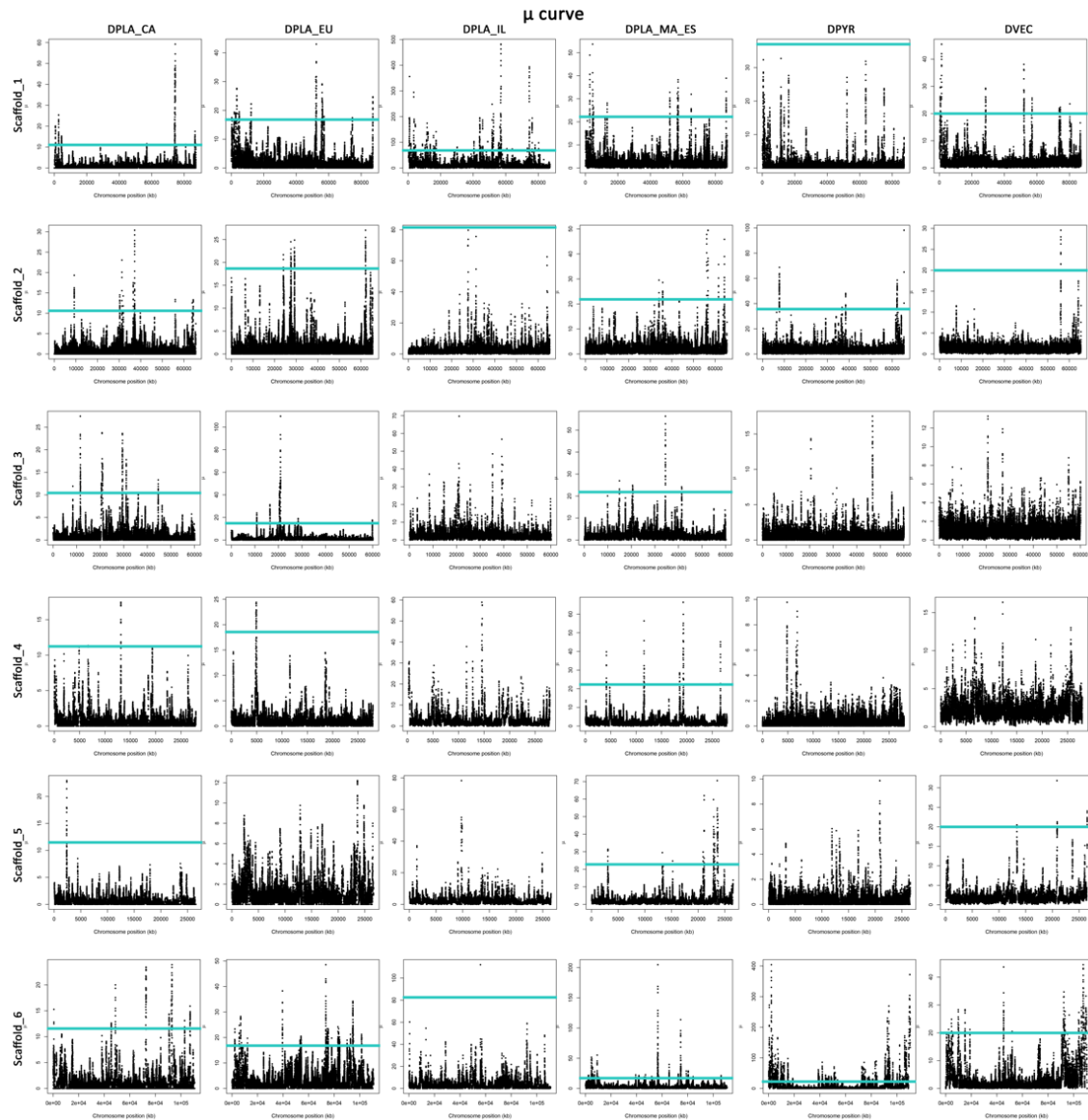


Figure 3. Signatures of positive selection in the six *Dysaphis* populations were inferred from the μ statistics estimated with RAiSD (Alanchiotis and Pavlidis, 2018; 7,622,598 SNPs). American group: DPLA_CA; European countries: DPLA_EU; Israeli group: DPLA_IL; *Dysaphis plantaginea* from Morocco and Spain: DPLA_MA_ES; *Dysaphis pyri*: DPYR; *Dysaphis devectora*: DVEC. The threshold value was defined as 1%, and all the SNP outliers are represented over the aqua line color

6.5 DISCUSSION AND PERSPECTIVES

In Chapter 2, we first provided insight into the *D. plantaginea* population genetic structure using microsatellite markers (Olvera-Vazquez et al., 2021). Here, we subsampled colonies representatives of this population genetic structure. We also included additional aphid individuals from North America and the Middle East to represent better the geographic distribution of *D. plantaginea* and added two sister species. We sequenced the samples using short-read Illumina technology. SNP markers combined with population genetics methods allow us to detect six main populations in our dataset. *Dysaphis devectora*, *D. pyri*, and *D. plantaginea* formed distinct populations. The clear pattern of genetic structure among the three species can result from demography and/or selection. The population genetic structure can reflect the different life cycles, heteroecious (*D. plantaginea* and *D. pyri*) and monoecious cycle (*D. devectora*) or hosts affinity *D. plantaginea* and *D. devectora* the cultivated apple and *D. pyri* cultivated pear, that lead to barrier to gene flow among species.

We could find support for this finding of the genetic structure if we detect genomic islands of differentiation among species using genome scans. My preliminary results using the μ statistics may indicate various islands of differentiation, but I need to dig into the details of those results to draw further conclusions. Such genomic islands of divergences have already been observed, for instance, between two ecotypes of the phytophagous insect *Timema cristinae* Vickery (Soria-Carrasco et al., 2014) found on two different plant hosts, *Adenostoma fasciculatum* Hook. & Arn. and *Ceanothus spinosus* Nutt. We will also further understand the process shaping the divergence of the three species by analyzing their speciation history and degree of gene flow among species.

Within *D. plantaginea*, we found weak genetic differentiation among the *D. plantaginea* populations. We observed a lower genetic diversity in the Israeli, North American, and Moroccan/Spanish populations, which suggests these populations are from recent colonization events. The Middle East may not be the center of origin of *D. plantaginea*. Europe seems to be a better candidate. This colonization scenario, however, needs to be further validated by investigating the demographic history of the species. We found a recent population expansion and split among the *D. plantaginea* genetic groups. However, to confirm this information we should try a complementary software because underestimations could occur with SMC++ with very recent demographic changes (Pattons et al., 2019; Wang et al; 2023). We will dig into the demography history of *D. plantaginea* using MSMC2 (Schiffels and Durbin, 2014). We will also add *D. pyri* and *D. devectora* to test their role in the colonization history of *D. plantaginea*. Indeed, diversity estimates showed that *D. pyri* had the highest genetic diversity and was the closest to *D. plantaginea*, suggesting a host jump from the pear tree in Europe. Another support for this hypothesis was the detection of the recent splitting from these two species but a more recent splitting from *D. pyri* than *D. devectora*. Moreover, the gene flow found between *D. plantaginea* groups and *D. pyri*. However, further inferences of the demographic history are needed (such as ABC-RF inferences). We will also estimate the degree of gene flow among populations with the software D-suite (Malinsky et al., 2021). Our search for selection signals in the four populations will also allow us to determine the local selective pressures related to abiotic and biotic factors that could have shaped the architecture of aphid genomes (Foll and Gaggiotti, 2006). Still, I need to dig into those results to draw further conclusions. To that end, I will compare our candidate SNPs found using the two statistics presented in Materials and Methods. I will also scan the genomes for signatures of balancing selection (Fijarczyk and Babik, 2015). I will also associate SNP

allele frequency with climate variables to detect candidate loci related to local adaptation to the abiotic environment (Duruz et al., 2019). These analyses would bring insights into key loci involved in the adaptation of aphids to the local biotic environment (*i.e.*, host) and abiotic environment. The analyses and results in this section will be complemented with the ecological patterns of local adaptation (Chapter 4).

6.6 CONCLUSION

Here, we provided preliminary insights into the role of selection and demography in the divergence of populations and species belonging to the *Dysaphis* genus (Blackman and Eastop, 2000). This genus includes significant fruit pest trees belonging to the Rosaceae family. We focus on two major crop pests, *D. plantaginea*, the rosy apple aphid, and *D. pyri*, or pear-bedstraw aphid, and one minor apple pest, the rosy leaf-curling apple aphid, *D. devectora*. We provided insights into the population genetic structure and candidate genes under positive selection in these species. These loci may be involved in the colonization success and adaptation of these pests to their host or local environments. However, I must dig into the results to draw further conclusions. This study represents an excellent opportunity to get an insight into the genes involved in speciation and colonization success in aphids, which is also relevant to preventing an outbreak in the cultivated apple due to the impact of global warming (Kirk et al., 2013; Pélissié et al., 2018).

6.8 REFERENCES

Alachiotis, N., & Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications biology*, 1(1), 79. <https://doi.org/10.1038/s42003-018-0085-8>

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Atlilan, R., Kasap, İ., Özgökçe, M. S., Polat-Akköprü, E., & Chi, H. (2017). Population growth of *Dysaphis pyri* (Hemiptera: Aphididae) on different pear cultivars with discussion on curve fitting in life table studies. *Journal of Economic Entomology*, 110(4), 1890-1898. <https://doi.org/10.1093/jee/tox174>

Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguilera, G., Snirc, A., ... & Giraud, T. (2017). Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Molecular Ecology*, 26(7), 2041-2062. <https://doi.org/10.1111/mec.13976>

Berlocher, S. H., & Feder, J. L. (2002). Sympatric speciation in phytophagous insects: moving beyond controversy?. *Annual review of entomology*, 47(1), 773-815. <https://doi.org/10.1146/annurev.ento.47.091201.145312>

Blackman, R. L., & Eastop, V. F. (2000). *Aphids on the world's crops: an identification and information guide* (No. Ed. 2). John Wiley & Sons Ltd

Brandt, D. Y., César, J., Goudet, J., & Meyer, D. (2018). The effect of balancing selection on population differentiation: a study with HLA genes. *G3: Genes, Genomes, Genetics*, 8(8), 2805-2815. <https://doi.org/10.1534/g3.118.200367>

Burke, M. K. (2012). How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749), 5029-5038. <https://doi.org/10.1098/rspb.2012.0799>

Calla, B., Demkovich, M., Siegel, J. P., Viana, J. P. G., Walden, K. K., Robertson, H. M., & Berenbaum, M. R. (2021). Selective sweeps in a nutshell: the genomic footprint of rapid insecticide resistance evolution in the almond agroecosystem. *Genome Biology and Evolution*, 13(1), evaa234. <https://doi.org/10.1093/gbe/evaa234>

Campos, J. L., & Charlesworth, B. (2019). The effects on neutral variability of recurrent selective sweeps and background selection. *Genetics*, 212(1), 287-303. <https://doi.org/10.1534/genetics.119.301951>

Cao, L. J., Song, W., Chen, J. C., Fan, X. L., Hoffmann, A. A., & Wei, S. J. (2022). Population genomic signatures of the oriental fruit moth related to the Pleistocene climates. *Communications Biology*, 5(1), 142. <https://doi.org/10.1038/s42003-022-03097-2>

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks : an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124-3140. <https://doi.org/10.1111/mec.12354>

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1). <https://doi.org/10.1186/s13742-015-0047-8>

Charlesworth, D., Charlesworth, B., & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics*, 141(4), 1619-1632. <https://doi.org/10.1093/genetics/141.4.1619>

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS genetics*, 2(4), e64. <https://doi.org/10.1371/journal.pgen.0020064>

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884-i890. <https://doi.org/10.1093/bioinformatics/bty560>

Chen, Y., Liu, Z., Régnière, J., Vasseur, L., Lin, J., Huang, S., ... & You, S. (2021). Large-scale genome-wide study reveals climate adaptive variability in a cosmopolitan pest. *Nature Communications*, 12(1), 7206. <https://doi.org/10.1038/s41467-021-27510-2>

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23), 3317-3324. <https://doi.org/10.1093/bioinformatics/btu530>

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of

single nucleotide polymorphisms, *SnEff. Fly*, 6(2), 80-92.
<https://doi.org/10.4161/fly.19695>

Crossley, M. S., Chen, Y. H., Groves, R. L., & Schoville, S. D. (2017). Landscape genomics of Colorado potato beetle provides evidence of polygenic adaptation to insecticides. *Molecular Ecology*, 26(22), 6284-6300.
<https://doi.org/10.1111/mec.14339>

Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, 14(4), 262-274.
<https://doi.org/10.1038/nrg3425>

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158.
<https://doi.org/10.1093/bioinformatics/btr330>

Drès, M., & Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1420), 471-492.
<https://doi.org/10.1098/rstb.2002.1059>

Duruz, S., Sevane, N., Selmoni, O., Vajana, E., Leempoel, K., Stucki, S., ... & Joost, S. (2019). Rapid identification and interpretation of gene-environment associations using the new R. SamBada landscape genomics pipeline. *Molecular ecology resources*, 19(5), 1355-1365. <https://doi.org/10.1111/1755-0998.13044>

Ebert, D., & Fields, P. D. (2020). Host-parasite co-evolution and its genomic signature. *Nature Reviews Genetics*, 21(12), 754-768. <https://doi.org/10.1038/s41576-020-0269-1>

Evans, K. M., Govan, C. L., & Fernandez-Fernandez, F. (2008). A new gene for resistance to *Dysaphis pyri* in pear and identification of flanking microsatellite markers. *Genome*, 51(12), 1026-1031. <https://doi.org/10.1139/G08-093>

Ferreira, E. A., Lambert, S., Verrier, T., Marion-Poll, F., & Yassin, A. (2020). Soft selective sweep on chemosensory genes correlates with ancestral preference for toxic noni in a specialist drosophila population. *Genes*, 12(1), 32.
<https://doi.org/10.3390/genes12010032>

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: limits and prospects. *Molecular ecology*, 24(14), 3529-3545. <https://doi.org/10.1111/mec.13226>

Fiteni, E., Durand, K., Gimenez, S., Meagher, R. L., Legeai, F., Kergoat, G. J., ... & Nam, K. (2022). Host-plant adaptation as a driver of incipient speciation in the fall armyworm (*Spodoptera frugiperda*). *BMC Ecology and Evolution*, 22(1), 1-11. <https://doi.org/10.1186/s12862-022-02090-x>

Francis, R.M. (2017). Pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources* 17, 27–32. <https://doi.org/10.1111/1755-0998.12509>

Foll, M., & Gaggiotti, O. (2006). Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, 174(2), 875-891 <https://doi.org/10.1534/genetics.106.059451>

Forrest, J. M. S., & Dixon, A. F. G. (1975). The induction of leaf-roll galls by the apple aphids *Dysaphis devecta* and *D. plantaginea*. *Annals of Applied Biology*, 81(3), 281-288. <https://doi.org/10.1111/j.1744-7348.1975.tb01643.x>

Gratwick, M. (1992). Apple aphids. In: Gratwick, M. (eds) *Crop Pests in the UK*. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-1490-5_3

Guillemaud, T., Blin, A., Simon, S., Morel, K., & Franck, P. (2011). Weak spatial and temporal population genetic structure in the rosy apple aphid, *Dysaphis plantaginea*, in French apple orchards. *Plos one*, 6(6), e21263. <https://doi.org/10.1371/journal.pone.0021263>

Hahn, M. W. (2019). *Molecular population genetics*. Sunderland, MA: Sinauer Associates. ISBN 978-0878939657

Hermisson, J., & Pennings, P. S. (2017). Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, 8(6), 700-716 <https://doi.org/10.1111/2041-210X.12808>

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., ... & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4), 379-397. <https://doi.org/10.1086/688018>

Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, 18(2), 337-338. <https://doi.org/10.1093/bioinformatics/18.2.337>

Huson, D. H., & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23(2), 254-267. <https://doi.org/10.1093/molbev/msj030>

Jensen, J. D. (2014). On the unfounded enthusiasm for soft selective sweeps. *Nature communications*, 5(1), 5281. <https://doi.org/10.1038/ncomms6281>

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, 15(5), 1179-1191. <https://doi.org/10.1111/1755-0998.12387>

Legeai, F., Shigenobu, S., Gauthier, J. P., Colbourne, J., Rispe, C., Collin, O., ... & Tagu, D. (2010). AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome. *Insect molecular biology*, 19, 5-12. <https://doi.org/10.1111/j.1365-2583.2009.00930.x>

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic acids research*, 49(W1), W293-W296. <https://doi.org/10.1093/nar/gkab301>

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987-2993. <https://doi.org/10.1093/bioinformatics/btr509>

Li, J., Li, H., Jakobsson, M., Li, S. E. N., Sjödin, P. E. R., & Lascoux, M. (2012). Joint analysis of demography and selection in population genetics: where do we stand and where could we go?. *Molecular ecology*, 21(1), 28-44. <https://doi.org/10.1111/j.1365-294X.2011.05308.x>

Li, Y. L., & Liu, J. X. (2017). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, 18(1), 176-177. <https://doi.org/10.1111/1755-0998.12719>

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Molecular ecology resources*, 21(2), 584-595. <https://doi.org/10.1111/1755-0998.13265>

Märkle, H., & Tellier, A. (2020). Inference of coevolutionary dynamics and parameters from host and parasite polymorphism data of repeated experiments. *PLoS Computational Biology*, 16(3), e1007668. <https://doi.org/10.1371/journal.pcbi.1007668>

Manel, S., Conord, C., & Després, L. (2009). Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. *BMC Evolutionary Biology*, 9(1), 1-10. <https://doi.org/10.1186/1471-2148-9-288>

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in ecology & evolution*, 28(11), 659-669. <https://doi.org/10.1016/j.tree.2013.08.003>

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37(5), 1530-1534. <https://doi.org/10.1093/molbev/msaa015>

Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39, 197-218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>

Nooten, S. S., & Hughes, L. (2017). The power of the transplant: direct assessment of climate change impacts. *Climatic Change*, 144, 237-255. <https://doi.org/10.1007/s10584-017-2037-6>

Olvera-Vazquez, S. G., Remoué, C., Venon, A., Rousselet, A., Grandcolas, O., Azrine, M., ... & Cornille, A. (2021). Large-scale geography survey provides insights into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa. *Peer Community Journal*, 1. <https://doi.org/10.24072/pcjournal.26>

Patton, A. H., Margres, M. J., Stahlke, A. R., Hendricks, S., Lewallen, K., Hamede, R. K., ... & Storfer, A. (2019). Contemporary demographic reconstruction methods are robust to genome assembly quality: a case study in Tasmanian devils. *Molecular biology and evolution*, 36(12), 2906-2921. <https://doi.org/10.1093/molbev/msz191>

Peischl, S., Dupanloup, I., Kirkpatrick, M., & Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions. *Molecular ecology*, 22(24), 5972-5982. <https://doi.org/10.1111/mec.12524>

Phipps, J. B., Robertson, K. R., Smith, P. G., & Rohrer, J. R. (1990). A checklist of the subfamily Maloideae (Rosaceae). *Canadian journal of botany*, 68(10), 2209-2269. <https://doi.org/10.1139/b90-288>

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3), 559-575. <https://doi.org/10.1086/519795>

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573-589. <https://doi.org/10.1534/genetics.114.164350>

Roche, P., Alston, F. H., Maliepaard, C., Evans, K. M., Vrieling, R., Dunemann, F., ... & King, G. J. (1997). RFLP and RAPD markers linked to the rosy leaf curling aphid resistance gene (Sd1) in apple. *Theoretical and Applied Genetics*, 94, 528-533. <https://doi.org/10.1007/s001220050447>

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature genetics*, 46(8), 919-925. <https://doi.org/10.1038/ng.3015>

Schiffels, S., & Wang, K. (2020). MSMC and MSMC2: the multiple sequentially markovian coalescent. In *Statistical population genomics* (pp. 147-165). Humana.

Siewert, K. M., & Voight, B. F. (2017). Detecting long-term balancing selection using allele frequency correlation. *Molecular biology and evolution*, 34(11), 2996-3005. <https://doi.org/10.1093/molbev/msx209>

Simon, J. C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquiere, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in functional genomics*, 14(6), 413-423. <https://doi.org/10.1093/bfpg/elv015>

Smadja, C. M., Canbäck, B., Vitalis, R., Gautier, M., Ferrari, J., Zhou, J. J., & Butlin, R. K. (2012). Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution*, 66(9), 2723-2738. <https://doi.org/10.1111/j.1558-5646.2012.01612.x>

Song, S. V., Downes, S., Parker, T., Oakeshott, J. G., & Robin, C. (2015). High nucleotide diversity and limited linkage disequilibrium in *Helicoverpa armigera* facilitates the detection of a selective sweep. *Heredity*, 115(5), 460-470. <https://doi.org/10.1038/hdy.2015.53>

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., ... & Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344(6185), 738-742. <https://doi.org/10.1126/science.1252136>

Stukenbrock, E. H., & Bataillon, T. (2012). A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. <https://doi.org/10.1371/journal.ppat.1002893>

Swofford, D. L., & Sullivan, J. (2003). Phylogeny inference based on parsimony and other methods using PAUP*. *The phylogenetic handbook: a practical approach to DNA and protein phylogeny, cáp*, 7, 160-206.

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature genetics*, 49(2), 303-309. <https://doi.org/10.1038/ng.3748>

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12), 673-680. <https://doi.org/10.1016/j.tree.2014.10.004>

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... & Banks, E. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics*, 43(1), 11-10. <https://doi.org/10.1002/0471250953.bi1110s43>

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, 106(supplement_1), 9939-9946. <https://doi.org/10.1073/pnas.0901397106>

Wang, X., Peischl, S., & Heckel, G. (2023). Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal. *Current biology*, 33(10), 2051-2062. <https://doi.org/10.1016/j.cub.2023.04.042>

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>

Wickham, H., François, R., Henry, L., Müller, K. (2022). *dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>

Wilke, C. O. (2020). *Cowplot: Streamlined Plot Theme and Plot Annotations for "Ggplot2"*. <https://CRAN.R-project.org/package=cowplot>

Yadav, S., Stow, A. J., & Dudaniec, R. Y. (2019). Detection of environmental and morphological adaptation despite high landscape genetic connectivity in a pest grasshopper (*Phaulacridium vittatum*). *Molecular Ecology*, 28(14), 3395-3412. <https://doi.org/10.1111/mec.15146>

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., & Yang, T. L. (2018). PopLDdecay : a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, 35(10), 1786-1788. <https://doi.org/10.1093/bioinformatics/bty875>

Zhao, Q., Shi, L., He, W., Li, J., You, S., Chen, S., ... & You, M. (2022). Genomic Variations in the Tea Leafhopper Reveal the Basis of Its Adaptive Evolution. *Genomics, Proteomics & Bioinformatics*, 20(6), 1092-1105. <https://doi.org/10.1016/j.gpb.2022.05.011>

6.9 SUPPLEMENTARY TEXT

Supplementary text 1: Filtering pipeline to obtain the *Dysaphis devectora*, *Dysaphis plantaginea*, and *Dysaphis pyri* genome variant call format (VCF).

The SNPs were called following the best practices of the Genome Analysis Toolkit pipeline (GATK; Van der Auwera et al., 2013). We first removed the SNPs corresponding to *Buchnera aphidicola* endosymbiont and the aphid mitogenome. We combined the gVCF files of each of the 174 aphid individuals. The gVCF contains information about every position in the genome (existing variant or not). We used the GATK to get a VCF file. We then applied hard filtering step removing low-quality variants (*i.e.*, variants with a low read depth) using a Quality by Depth (QD) <2.0, Fisher Strand (FS) >60, Strand Odd Ratio (SOR) >3, RMS Mapping Quality (MQ) <40, Mapping Quality Rank Sum Test (MQRankSum) <- 12.5, resulting in 41,706,601 SNPs for the 174 aphid individuals. We had a checkpoint to identify the individuals to remove with a high proportion of missing data based on the read depth (DP) between DP < 5 and DP >100 and remove individuals with less than 30% of missing SNPs. We removed six individuals, keeping 168 individuals. We applied filtering of variants and invariants using BCFtools (Li et al., 2009) and VCFtools (Danecek et al., 2011), keeping 16,377,788 SNPs. We excluded SNPs with > 20% of missing genotypes and filtered for minor allele frequency (maf<0.05), keeping a total of 7,622,598 SNPs. We then sought clones using a kinship cut-off of 0.354 (Kelmemei et al., 2015) implemented in Plink software (Purcell et al., 2017), leaving 7,622,598 SNPs for 148 individuals. We applied a non-synonymous filtering step. We finally removed SNPs that were in Linkage Disequilibrium (LD) in a window size of 1Kb using VCFtools, which has a final genome VCF file of 49,964 SNPs with 148 individuals.

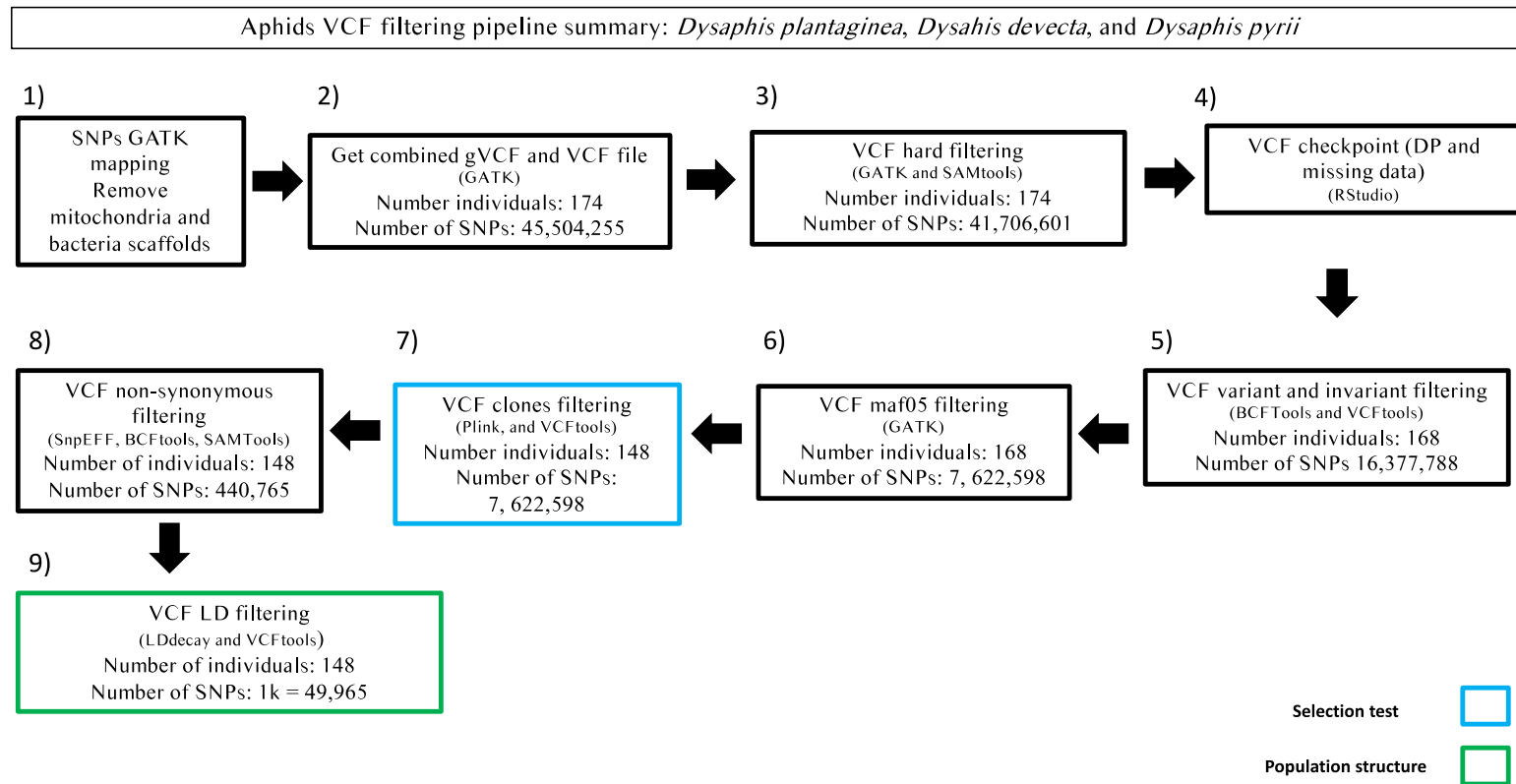


Figure 1. Filtering pipeline to obtain the *Dysaphis devectora*, *Dysaphis plantaginea*, and *Dysaphis pyrii* genome variant call format (VCF). 1) SNPs GATK mapping and removal of mitochondria and bacteria scaffolds. 2) Combination of all samples gVCF files and generation of VCF file using GATK, resulting in 174 individuals and 45,504,255 SNPs. 3) VCF hard filtering resulted in 174 individuals and 41,706,601 SNPs. 4) Checkpoint to detect the individuals based on read depth (DP) between $DP < 5$ and $DP > 100$ and individuals with less than 30% of missing SNPs. 5) Filtering of variant and invariant filtering step resulting in 168 individuals and 16,377,788 SNPs. 6) Filtering step removing $> 20\%$ of missing genotypes and filtered for minor allele frequency ($maf < 0.05$), keeping a total of 7,622,598 SNPs. 7) Clone detection and removing step resulting in 148 individuals and 7,622,598 SNPs. This file was used for selection test analysis. 8) Filtering for non-synonymous sites resulted in 148 individuals and 440,765 SNPs. 9) Filtering 1Kb windows to remove linkage disequilibrium (LD) sites, giving 148 individuals and leaving 49,965 SNPs. This file was utilized for evaluating the aphid population structure.

REFERENCES

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. <https://doi.org/10.1093/bioinformatics/btr330>

Kelmemi, W., Teeuw, M. E., Bochdanovits, Z., Ouburg, S., Jonker, M. A., AlKuraya, F., ... & Cornel, M. C. (2015). Determining the genome-wide kinship coefficient seems unhelpful in distinguishing consanguineous couples with a high versus low risk for adverse reproductive outcome. *BMC Medical Genetics*, 16, 1-11. <https://doi.org/10.1186/s12881-015-0191-0>

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *bioinformatics*, 25(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>

6.10 SUPPLEMENTARY MATERIAL

Table S1. Aphid general information showing the number of individuals, sequence ID, Data ID, Site ID, pest species name, hostname, country, locality, latitude, longitude, apple variety, sampler, information about the resequencing genome (library name, flowcell/lane, raw reads, raw data, percentage of effective reads, percentage of errors, percentage of quartile 20, percentage of quartile 30, percentage of GC content, mar reads, map base), and the information of the membership assignment K9 using FastStructure (Raj et al., 2014). Excel file.

Table S2. Summary of *D-statistic* of testing gene flow among *D. plantaginea* genetic groups (Hypothesis 1).

Table S3. Summary of *D-statistic* of testing gene flow between *D. plantaginea* and *D. pyri* genetic groups (Hypothesis 2).

Figure S1. Linkage disequilibrium decay using PopLDdecay (Zhang *et al.*, 2018) of the three aphid populations showing a windows distance of 300Kb (left) and increasing the windows to 10 Kb (right) the plots were built using R software v. 4.1.2 (R Core Team, 2021).

Figure S2. Samples used to scan the genomes of *Dysaphis* spp. for positive selection and pattern of adaptation and local adaptation. After filtering, we included 14 samples of *Dysaphis devectora* from Romania and Iran, 118 individuals of *Dysaphis plantaginea* from Canada, Belgium, Bulgaria, Denmark, France, Germany, Italy, Israel, Morocco, Romania, Sweden, Spain, UK, and the US, and 11 Individuals of *Dysaphis pyri* from Spain, France, and Iran.

Figure S3. Population genetic structure for *Dysaphis devectora* ($N = 118$), *Dysaphis plantaginea* ($N = 14$), and *Dysaphis pyri* ($N = 11$) inferred using fastSTRUCTURE with logarithm function (Raj et al., 2014) from $K=2$ to $K=20$. CA: Canada; US: the US; MA: Morocco; IL: Israel; ES: Spain; FR: France; DE: Germany; BE: Belgium; GB: Great Britain; SE: Sweden; DK: Denmark; BG: Bulgaria; IT: Italy; RO: Romania; Dv: *Dysaphis devectora*, Dpi: *Dysaphis pyri*.

Figure S4. Optimal K value assessed using StructureSelector with a Puechemaille threshold value of 0.5 (Li and Li, 2017) from fastSTRUCTURE outputs ran from $K=2$ to $K=20$ for *Dysaphis devectora* ($N = 118$), *Dysaphis plantaginea* ($N = 14$) and *Dysaphis pyri* ($N = 11$).

Figure S5. Membership coefficient of the *Dysaphis devectora*, *Dysaphis plantaginea*, and *Dysaphis pyri*. The plot of the membership coefficient of the three aphid species clusters, *D. devectora*, *D. pyri*, and *D. plantaginea*, using $K=19$ optimal value based on the structure selector, is presented on the left side of the figure. All the individuals present assignment values >90 ($N=143$). The plot of the membership coefficient only for *D. plantaginea* individuals using $K = 9$ optimal values based on the structure selector is presented on the right side of the figure. The number of individuals <90 was 90 ($N = 118$).

Figure S6. Principal component analysis (PCA) of the genetic variation of *Dysaphis plantaginea* ($N = 118$, blue), *Dysaphis devectora* ($N = 14$, green), and *Dysaphis pyri* ($N = 11$, red) samples (49,965 unlinked synonymous SNPs). The eigenvalues obtained with Plink v2 (Chang et al., 2015) were plotted using the R package ggplot2 (Wickham, 2009).

Figure S7. Genetic relationships between *Dysaphis plantaginea* ($N = 118$, blue), *Dysaphis devectora* ($N = 14$, green), and *Dysaphis pyri* ($N = 11$, red) based on the mitochondrial genomes (2,654 SNPs) built with A) with IQ-Tree (Minh et al., 2020) using maximum likelihood tree model finder option, ultrafast bootstrap 1,000, and 1,000 branch support tests (SH-aLRT), the tree was built using iTOL v5 (Letunic and Bork, 2021); B) Neighbour net using with Splistree v4 (Huson and Bryant, 2005).

Figure S8. Genetic relationships between *Dysaphis plantaginea* ($N = 118$, blue), *Dysaphis devectora* ($N = 14$, green), and *Dysaphis pyri* ($N = 11$, red) based on the *Buchnera aphidicola* genomes (121,311 SNPs) built with A) with IQ-Tree (Minh et al., 2020) using maximum likelihood tree model finder option, ultrafast bootstrap 1,000, and 1,000 branch support tests (SH-aLRT), the tree was built using iTOL v5 (Letunic and Bork, 2021); B) Neighbour net using with Splistree v4 (Huson and Bryant, 2005).

Figure S9. *Dysaphis plantaginea* genetic structure inferred with fastSTRUCTURE (Raj et al., 2014) with logarithm option standard option (left side of the figure $N = 118$) and from $K=2$ to $K=10$.

Figure S10. Analysis for detecting the optimal K value using StructureSelector with a threshold value of 0.5 (Li and Li, 2017) from $K=2$ to $K=10$ for *Dysaphis devectora* ($N = 118$).

Figure S11. Genetic relationships for *Dysaphis plantaginea* ($N = 118$) based on the mitochondrial genome (2,654 SNPs). A) Genetic relationships with IQ-Tree (Minh et al., 2020) using maximum likelihood tree model finder option, ultrafast bootstrap 1,000, and 1,000 branch support tests (SH-aLRT), the tree was built using iTOL v5 (Letunic and Bork, 2021), showing a clear group for America, Israel, and

Morocco. A lack of geographical genetic structure was detected among the different European countries. B) Neighbour net built with Splistree v4 (Huson and Bryant, 2005) among the *D. plantaginea* countries.

Figure S12. Genetic relationships for *Dysaphis plantaginea* samples ($N = 118$) based on the *Buchnera aphidicola* genome (121,311 SNPs). A) Genetic relationships with IQ-Tree (Minh et al., 2020) using maximum likelihood tree model finder option, ultrafast bootstrap 1,000, and 1,000 branch support tests (SH-aLRT), the tree was built using iTOL v5 (Letunic and Bork, 2021). B) Neighbour net built with Splistree v4 (Huson and Bryant, 2005) among the different *D. plantaginea* countries.

Figure S13. European *Dysaphis plantaginea* genetic structure using fastSTRUCTURE (Raj et al., 2014) logarithm option ($N = 77$) from $K=2$ to $K=20$.

Figure S14. European *Dysaphis plantaginea* genetic structure using fastSTRUCTURE (Raj et al., 2014) logarithm option ($N = 77$) from $K=2$ to $K=20$.

Figure S15. Demographic changes of N_e for four *D. plantaginea* genetic groups in the recent 10^8 years using SMC++ (Terhorst et al., 2017). The orange line corresponds to the Holocene period. N_America = *D. plantaginea* from Canada and the US; N_Israel = *D. plantaginea* from Israel; N_Europe = *D. plantaginea* from Belgium, Denmark, France, Germany, Romania, Spain, Sweden, and the UK; N_MA_ES = *D. plantaginea* from Morocco and Spain.

Figure S16. Demographic changes of N_e for A) *D. pyri* (N_DPYR) and B) *D. devecta* (D_DVEC) genetic groups in the recent 10^8 years using SMC++ (Terhorst et al., 2017). The orange line corresponds to the Holocene period.

Figure S17. Splitting time of the aphid population genetic groups inferred for A) *D. pyri* and B) *D. devectora* genetic groups in the recent 10^8 years using SMC++ (Terhorst et al., 2017). The blue line corresponds to the split time. The aphid genetic groups are: N_America = *D. plantaginea* from Canada and the US; N_Israel = *D. plantaginea* from Israel; N_Europe = *D. plantaginea* from Belgium, Denmark, France, Germany, Romania, Spain, Sweden, and the UK; N_MA_ES = *D. plantaginea* from Morocco and Spain; N_DPYR = *D. pyri*; N_DVEC = *D. devectora*.

Figure S18. Aphid species-tree inference with the single value decomposition quartets (SVDQuartets) score (Chifman and Kubatko, 2014). Bootstrap values are described in the branches of the tree.

Figure S19. Heatmap of *D-statistic* values (Malinsky et al., 2020) of the tested hypotheses: A) Gene flow among *D. plantaginea* genetic groups. B) Gene flow between *D. plantaginea* and *D. pyri* genetic groups. The lower right panel shows the absolute *D-statistic* value in red color the higher while the lower in blue color.

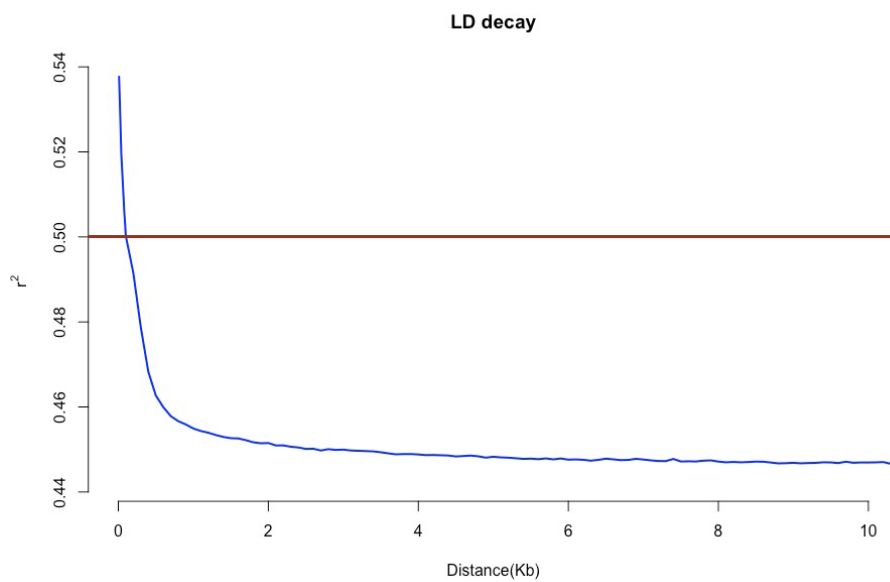
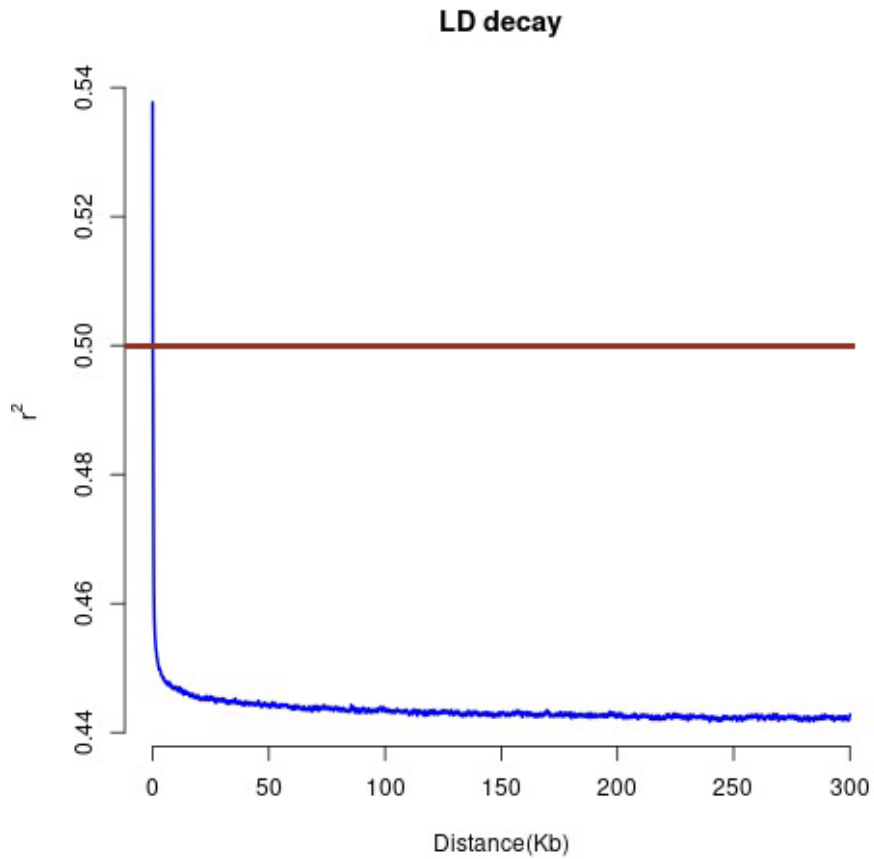


Figure S1. Linkage disequilibrium decay using PopLDdecay (Zhang *et al.*, 2018) of the three aphid populations showing a windows distance of 300Kb (up) and increasing the windows to 10 Kb (down) the plots were built using R software v. 4.1.2 (R Core Team, 2021).

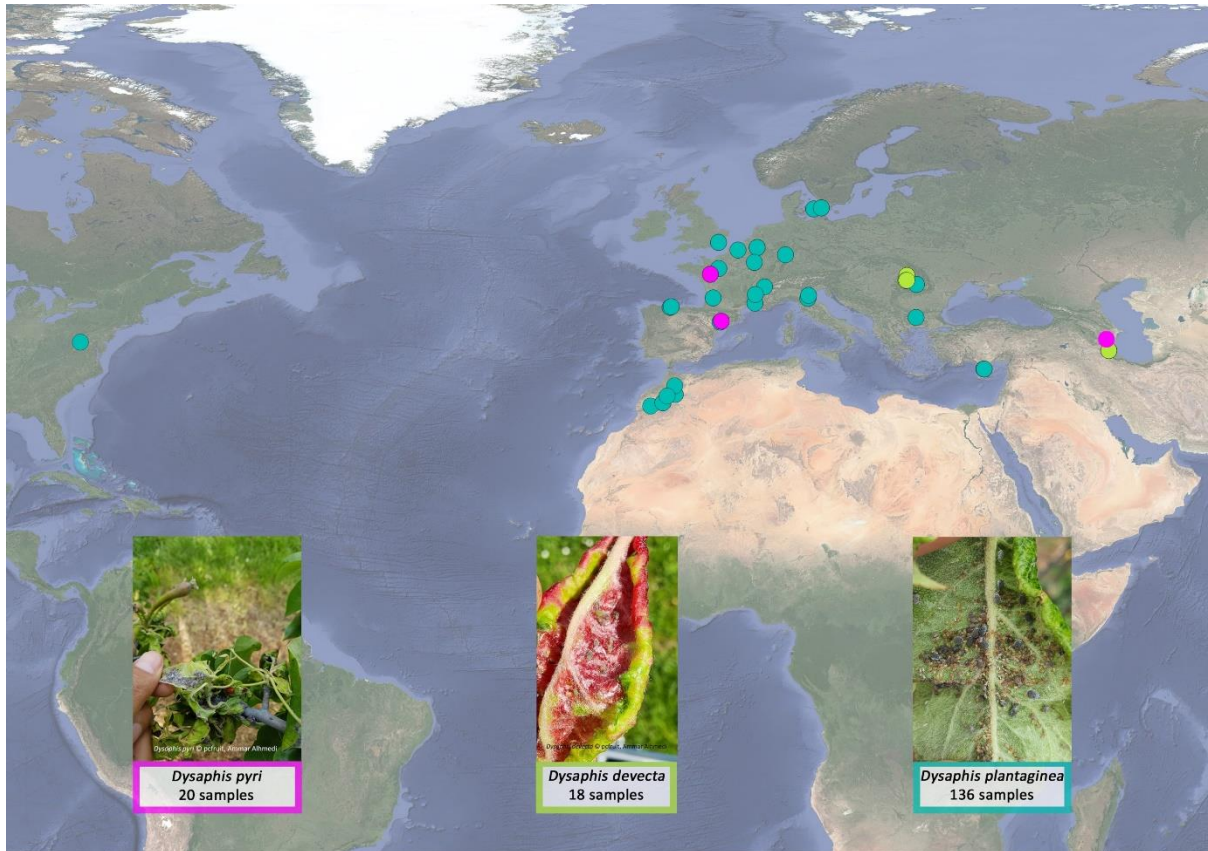


Figure S2. Samples used to scan the genomes of *Dysaphis* spp. for positive selection and pattern of adaptation and local adaptation. After filtering, we included 14 samples of *Dysaphis devectora* from Romania and Iran, 118 individuals of *Dysaphis plantaginea* from Canada, Belgium, Bulgaria, Denmark, France, Germany, Italy, Israel, Morocco, Romania, Sweden, Spain, UK, and the US, and 11 Individuals of *Dysaphis pyri* from Spain, France, and Iran.

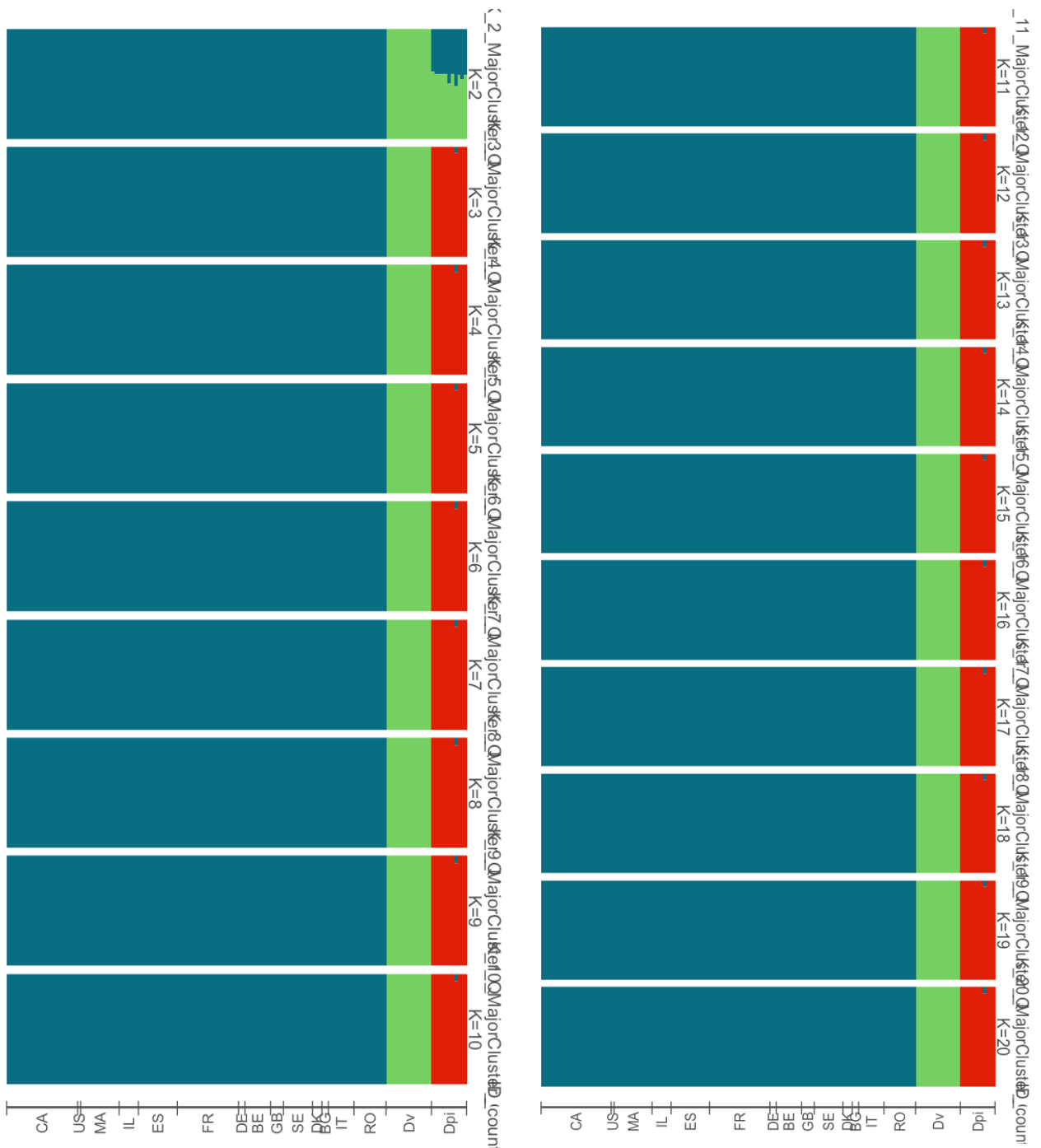


Figure S3. Population genetic structure for *Dysaphis devecta* ($N = 118$), *Dysaphis plantaginea* ($N = 14$), and *Dysaphis pyri* ($N = 11$) inferred using fastSTRUCTURE with logarithm function (Raj et al., 2014) from $K=2$ to $K=20$. CA: Canada; US: the US; MA: Morocco; IL: Israel; ES: Spain; FR: France; DE: Germany; BE: Belgium; GB: Great Britain; SE: Sweden; DK: Denmark; BG: Bulgaria; IT: Italy; RO: Romania; Dv: *Dysaphis devecta*, Dpi: *Dysaphis pyri*.

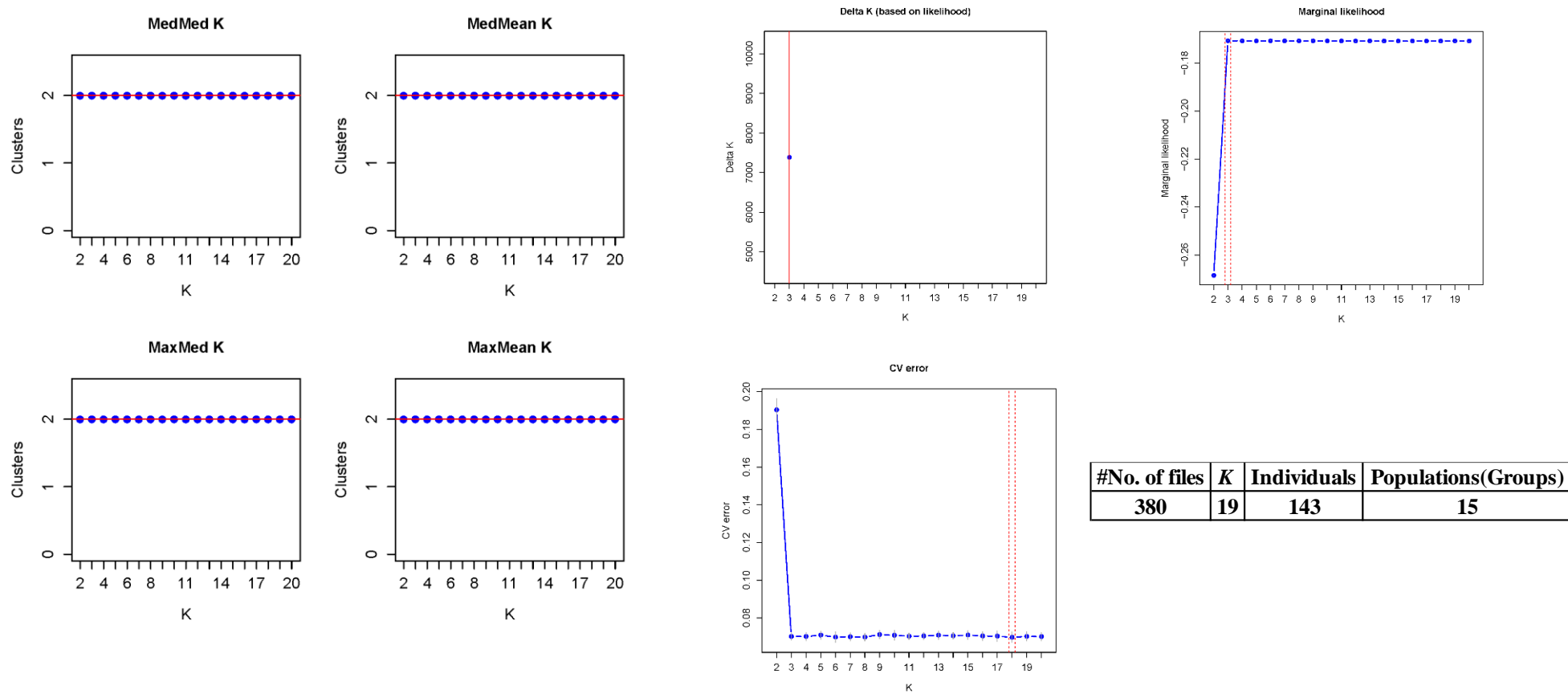


Figure S4. Optimal K value assessed using StructureSelector with a *Puechemaille* threshold value of 0.5 (Li and Li, 2017) from fastSTRUCTURE outputs ran from $K=2$ to $K=20$ for *Dysaphis plantaginea* ($N = 118$), *Dysaphis devectora* ($N = 14$) and *Dysaphis pyri* ($N = 11$).

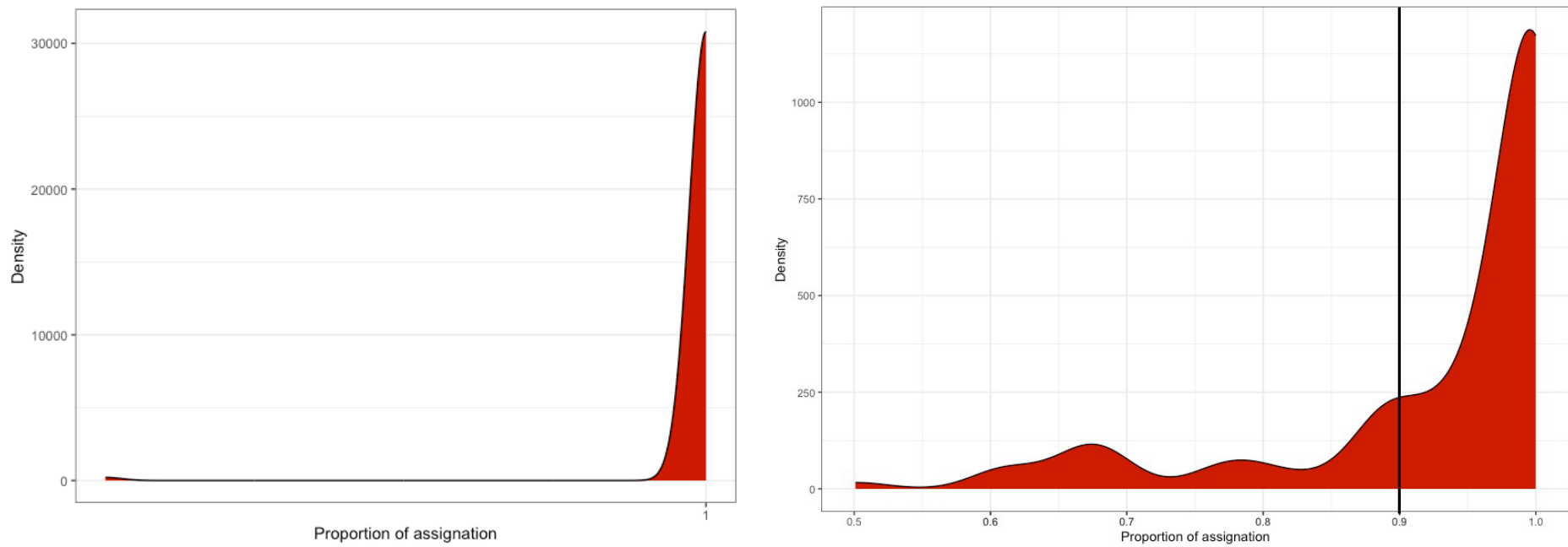


Figure S5. Distribution of the maximum membership coefficients inferred with fastSTRUCTURE for the full dataset (3 species, $N=143$, left figure), and *Dysaphis plantaginea* only ($N=118$, right figure). Distribution of the maximum membership coefficient of the three aphid species, *D. devecta*, *D. pyri*, and *D. plantaginea*, inferred at $K=3$. All the individuals present assignment values >90 ($N=143$). The distribution of the maximum membership coefficient for *D. plantaginea* only inferred at $K=9$. The number of individuals <90 was 90 ($N=118$).

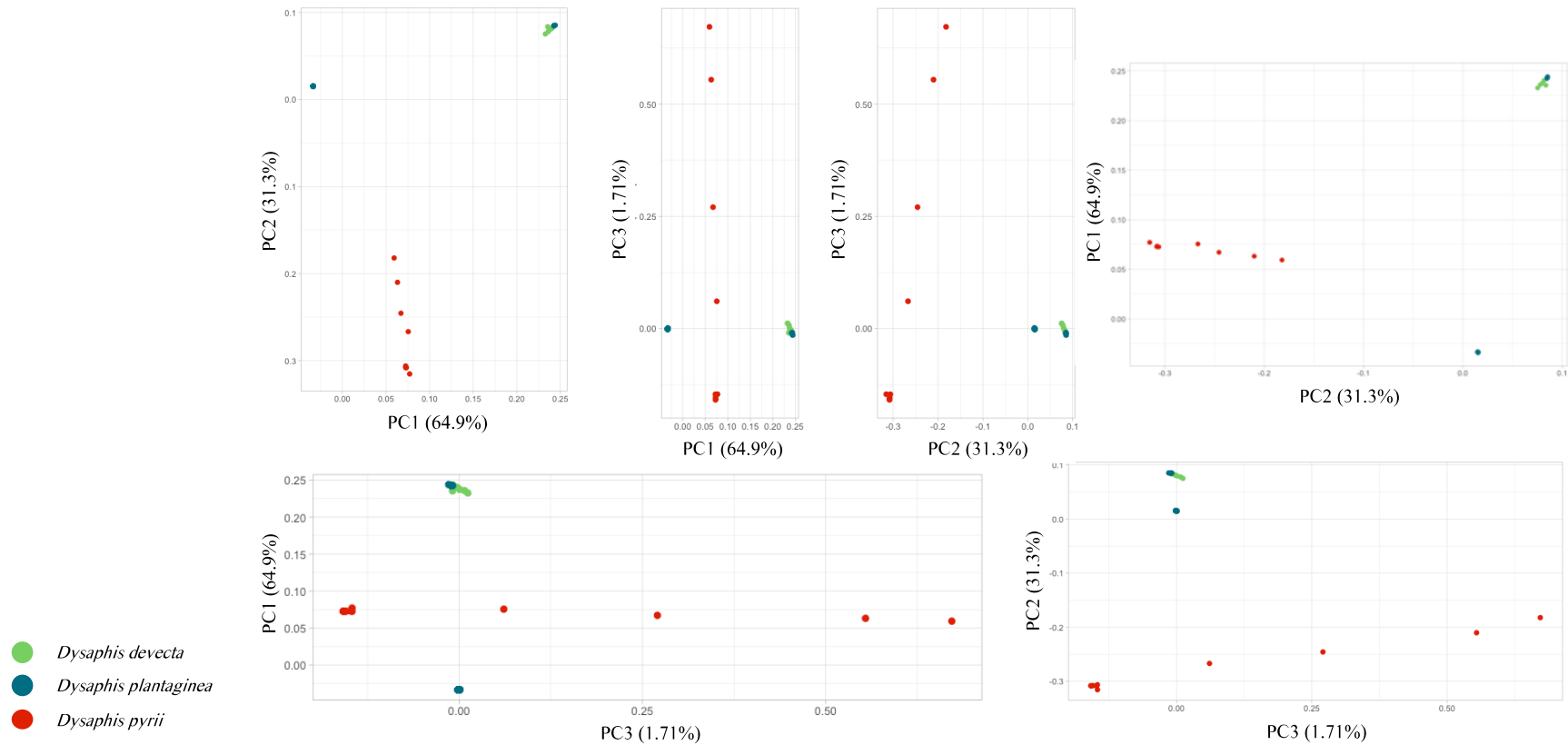


Figure S6. Principal component analysis (PCA) of the genetic variation of *Dysaphis plantaginea* ($N = 118$, blue), *Dysaphis devecta* ($N = 14$, green), and *Dysaphis pyri* ($N = 11$, red) samples (49,965 unlinked synonymous SNPs). The eigenvalues obtained with Plink v2 (Chang et al., 2015) were plotted using the R package ggplot2 (Wickham, 2009).

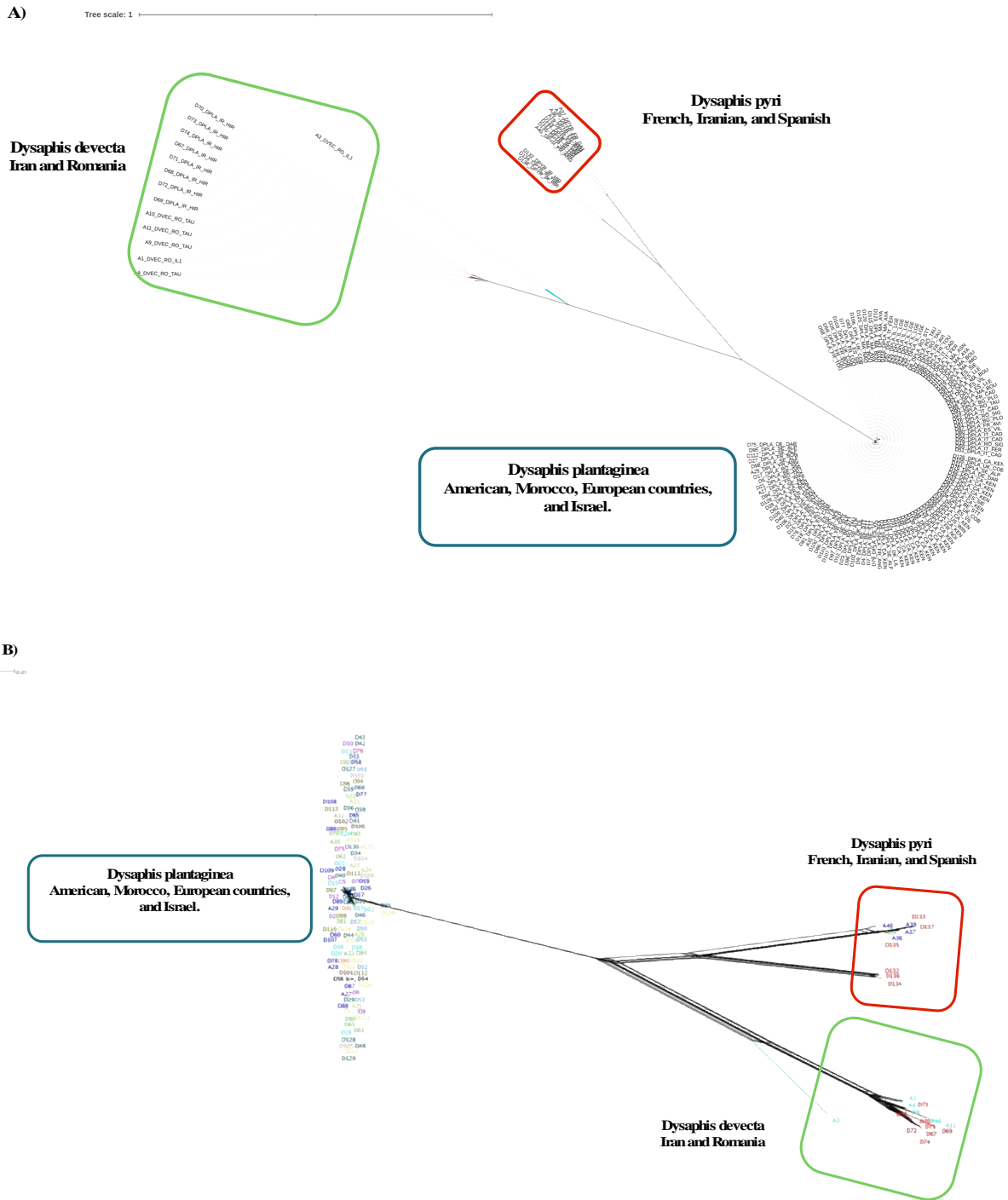


Figure S7. Genetic relationships between *Dysaphis plantaginea* ($N = 118$, blue), *Dysaphis devectora* ($N = 14$, green), and *Dysaphis pyri* ($N = 11$, red) based on the mitochondrial genomes (2,654 SNPs) built with A) with IQ-Tree (Minh et al., 2020) using maximum likelihood tree model finder option, ultra-fast bootstrap 1,000, and 1,000 branch support tests (SH-aLRT), the tree was built using iTOL v5 (Letunic and Bork, 2021); B) Neighbour net using with Splistree v4 (Huson and Bryant, 2005).

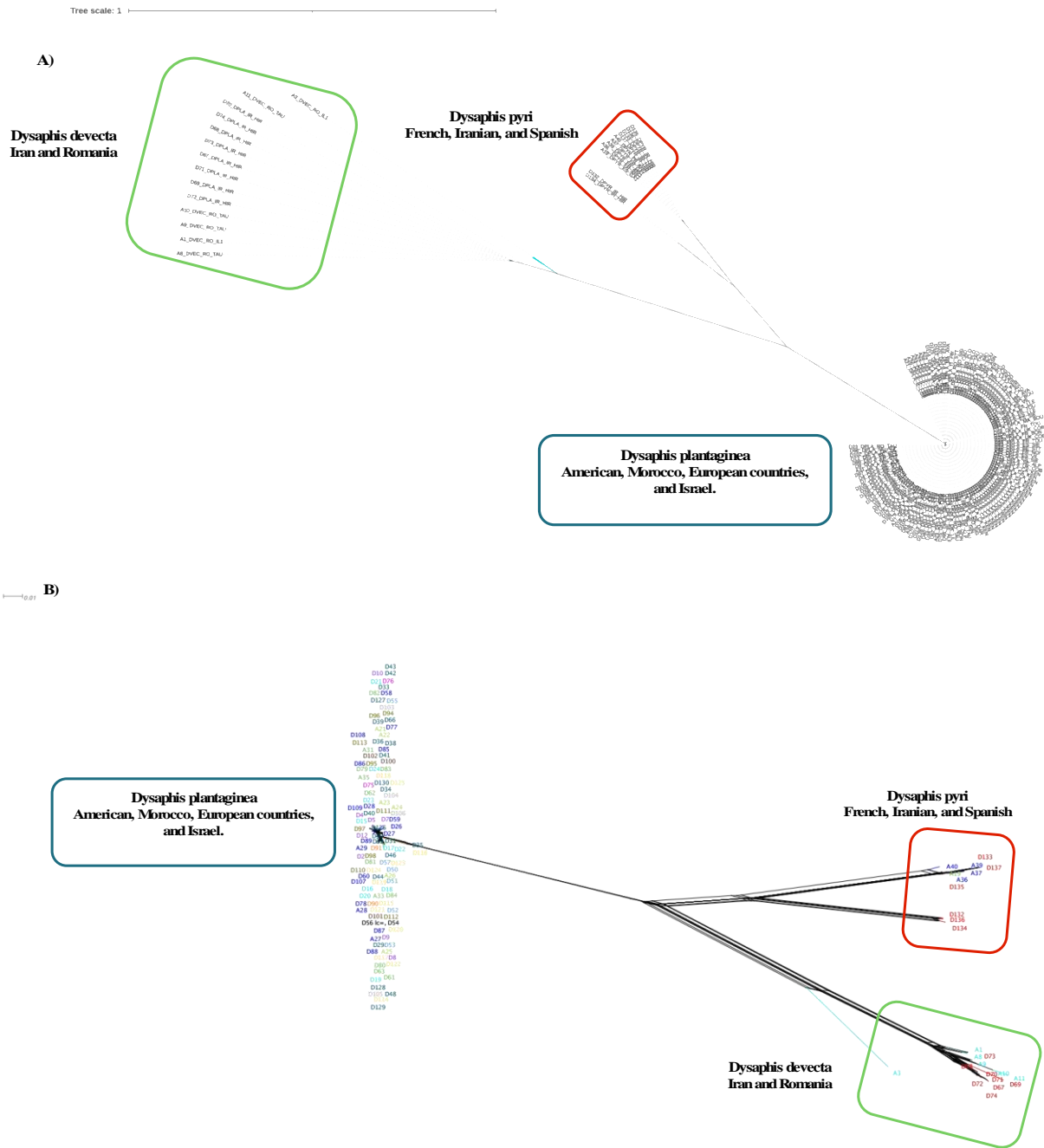


Figure S8. Genetic relationships between *Dysaphis plantaginea* ($N = 118$, blue), *Dysaphis devecta* ($N = 14$, green), and *Dysaphis pyri* ($N = 11$, red) based on the *Buchnera aphidicola* genomes (121,311 SNPs) built with A) with IQ-Tree (Minh et al., 2020) using maximum likelihood tree model finder option, ultrafast bootstrap 1,000, and 1,000 branch support tests (SH-aLRT), the tree was built using iTOL v5 (Letunic and Bork, 2021); B) Neighbour net using with Splistree v4 (Huson and Bryant, 2005).

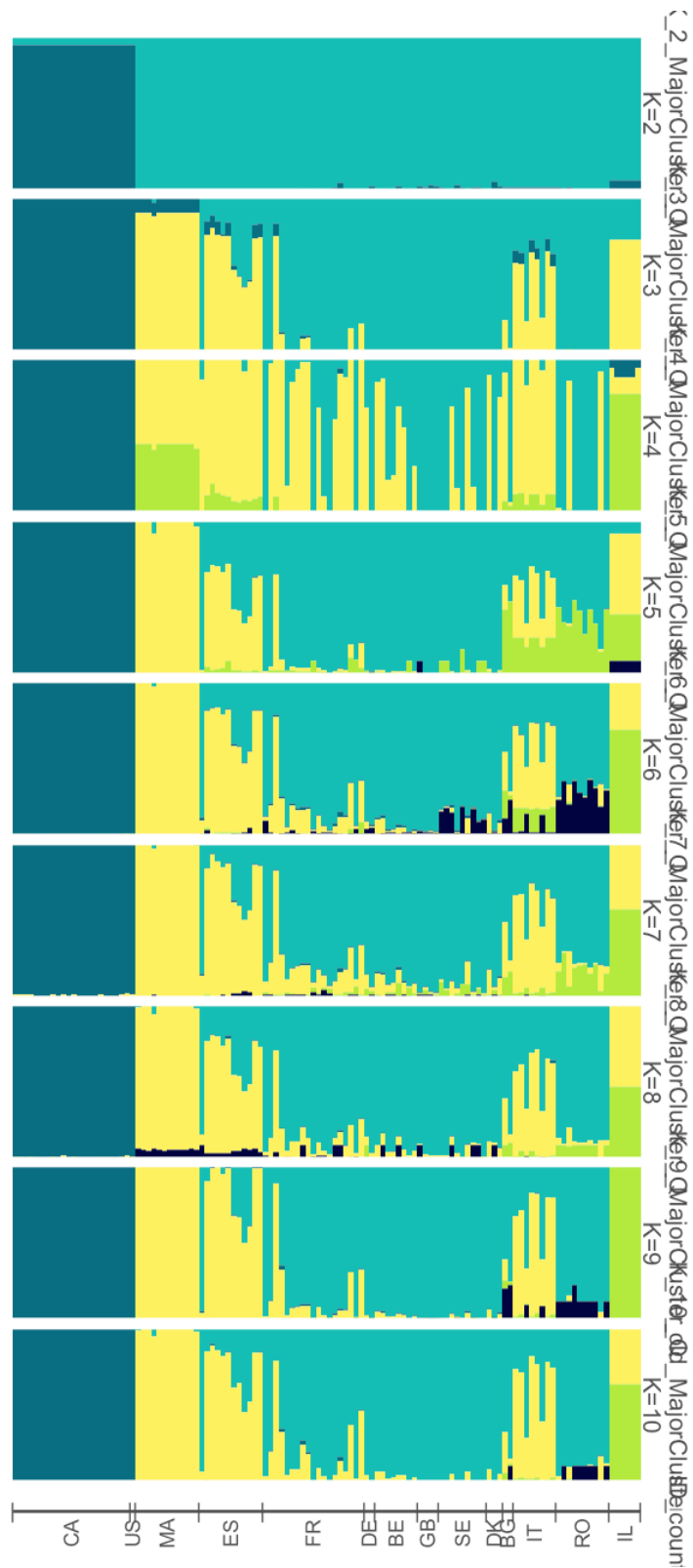
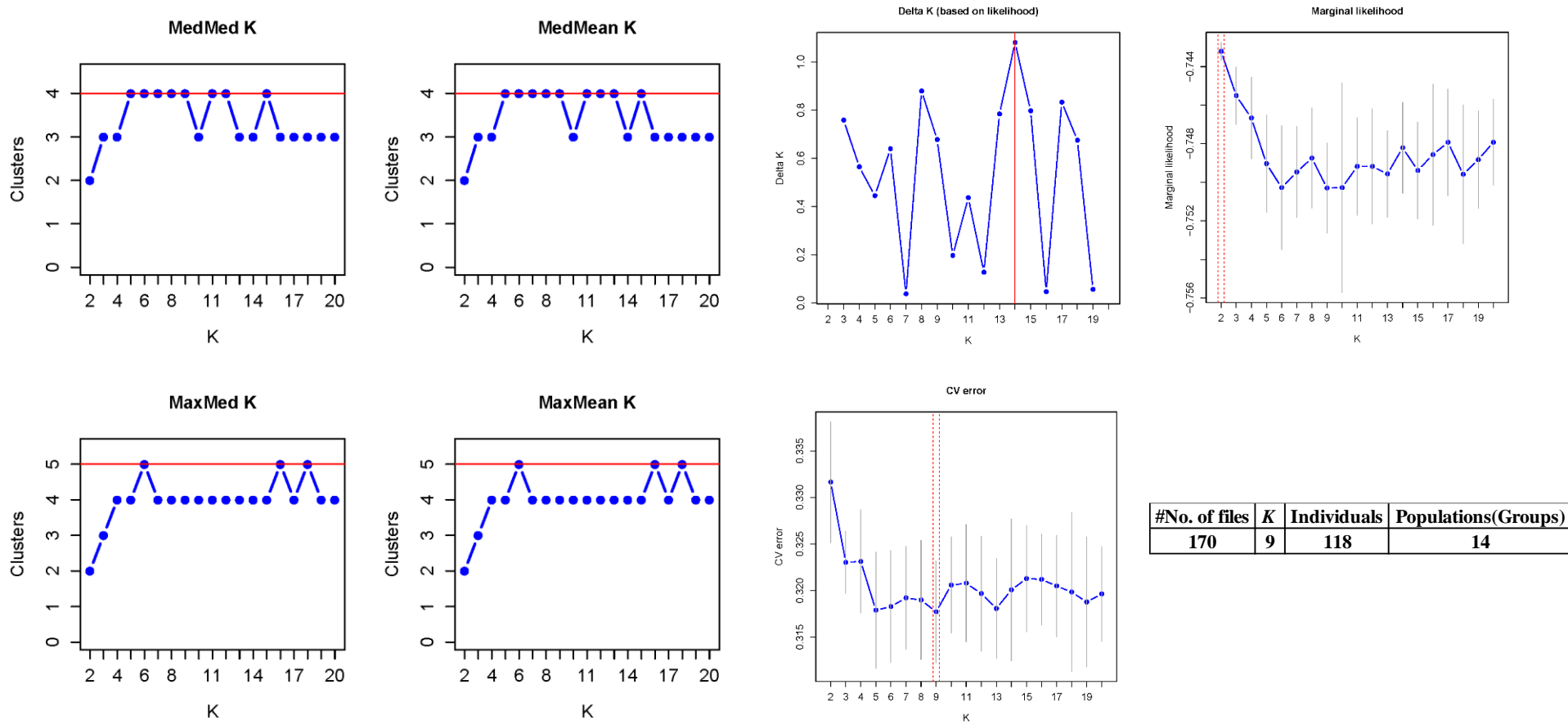
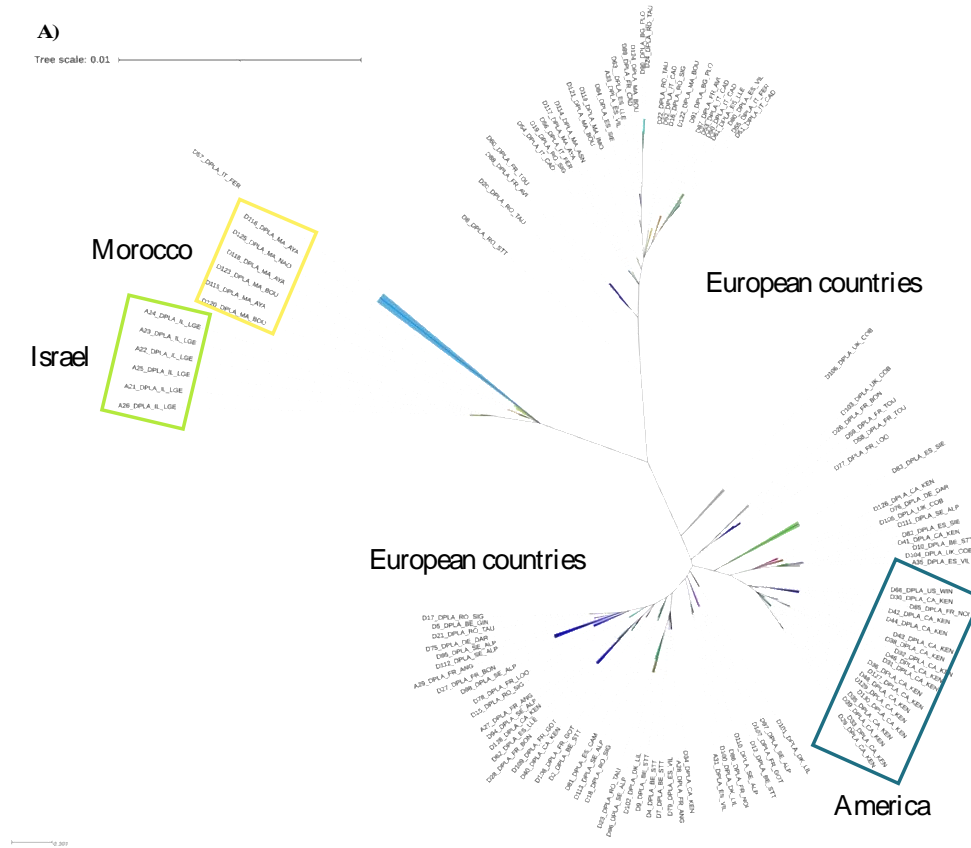


Figure S9. *Dysaphis plantaginea* genetic structure inferred with fastSTRUCTURE (Raj et al., 2014) with logarithm option standard option (left side of the figure $N = 118$) and from $K=2$ to $K=10$. CA: Canada; US: the US; MA: Morocco; IL: Israel; ES: Spain; FR: France; DE: Germany; BE: Belgium; GB: Great Britain; SE: Sweden; DK: Denmark; BG: Bulgaria; IT: Italy; RO: Romania.



#No. of files	K	Individuals	Populations(Groups)
170	9	118	14

Figure S10. Analysis for detecting the optimal K value using StructureSelector with a threshold value of 0.5 (Li and Li, 2017) from $K=2$ to $K=10$ for *Dysaphis devecta* ($N = 118$)



B)

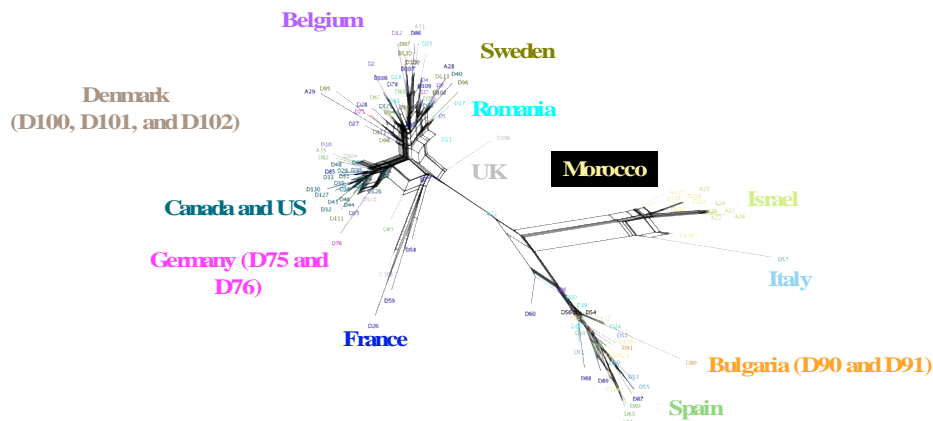


Figure S11. Genetic relationships for *Dysaphis plantaginea* ($N = 118$) based on the mitochondrial genome (2,654 SNPs). A) Genetic relationships with IQ-Tree (Minh et al., 2020) using maximum likelihood tree model finder option, ultrafast bootstrap 1,000, and 1,000 branch support tests (SH-aLRT), the tree was built using iTOL v5 (Letunic and Bork, 2021), showing a clear group for America, Israel, and Morocco. A lack of geographical genetic structure was detected among the different European countries. B) Neighbour net built with Splistree v4 (Huson and Bryant, 2005) among the *D. plantaginea* countries.

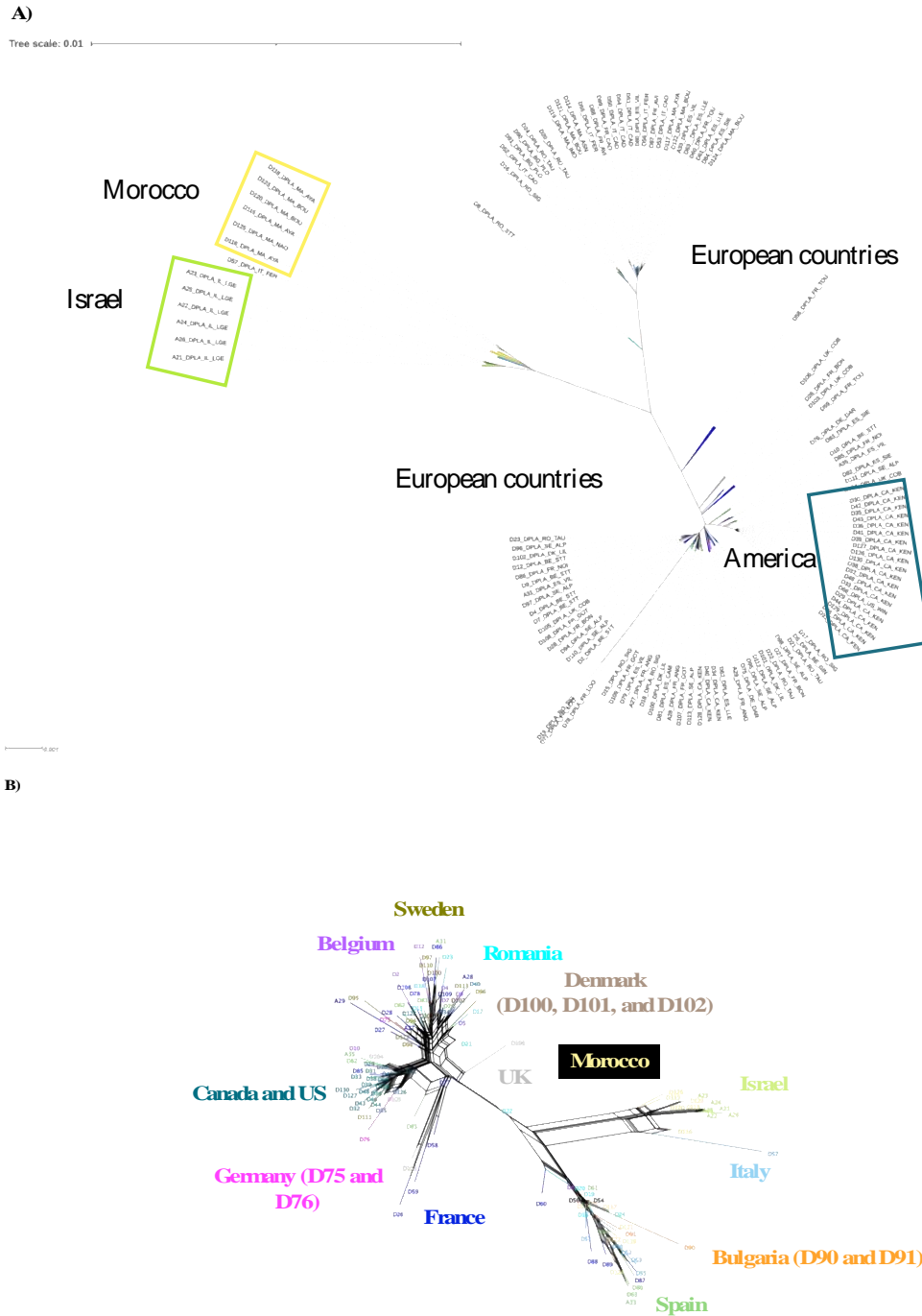


Figure S12. Genetic relationships for *Dysaphis plantaginea* samples ($N = 118$) based on the *Buchnera aphidicola* genome (121,311 SNPs). A) Genetic relationships with IQ-Tree (Minh et al., 2020) using maximum likelihood tree model finder option, ultrafast bootstrap 1,000, and 1,000 branch support tests (SH-aLRT), the tree was built using iTOL v5 (Letunic and Bork, 2021). B) Neighbour net built with Splistree v4 (Huson and Bryant, 2005) among the different *D. plantaginea* countries.

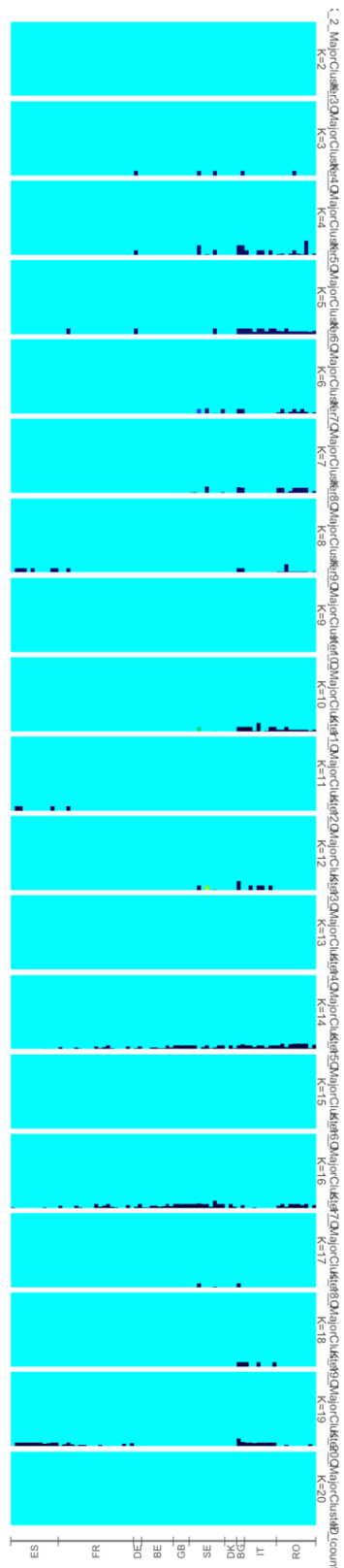


Figure S13. Population genetic structure of *Dysaphis plantaginea* of Europe only using fastSTRUCTURE (Raj et al., 2014) logarithm option ($N = 77$) from $K=2$ to $K=20$. ES: Spain; FR: France; DE: Germany; BE: Belgium; GB: Great Britain; SE: Sweden; DK: Denmark; BG: Bulgaria; IT: Italy; RO: Romania.

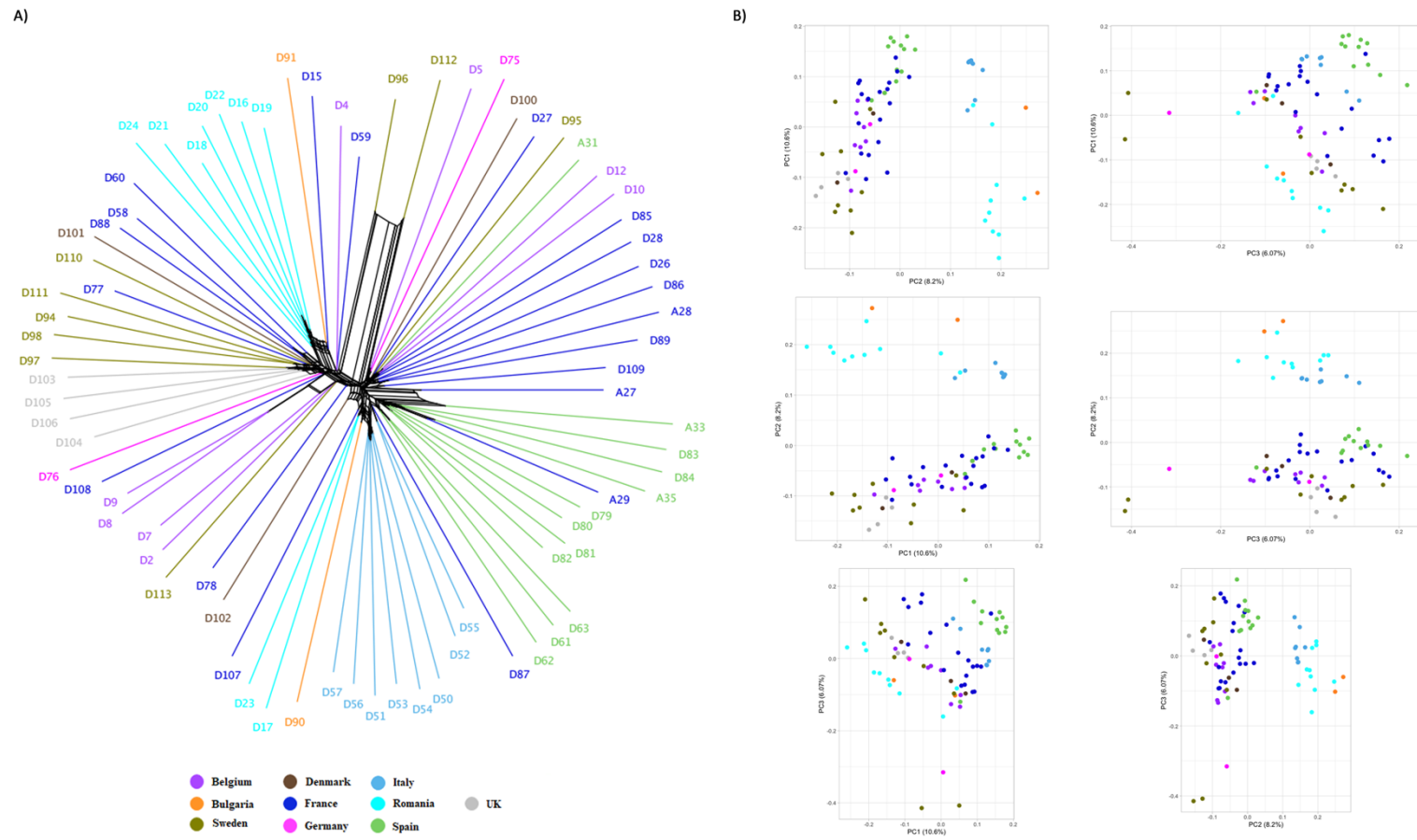


Figure S14. Principal component analyses of *Dysaphis plantaginea* samples from Europe ($N = 77$, 49,965 SNPs).

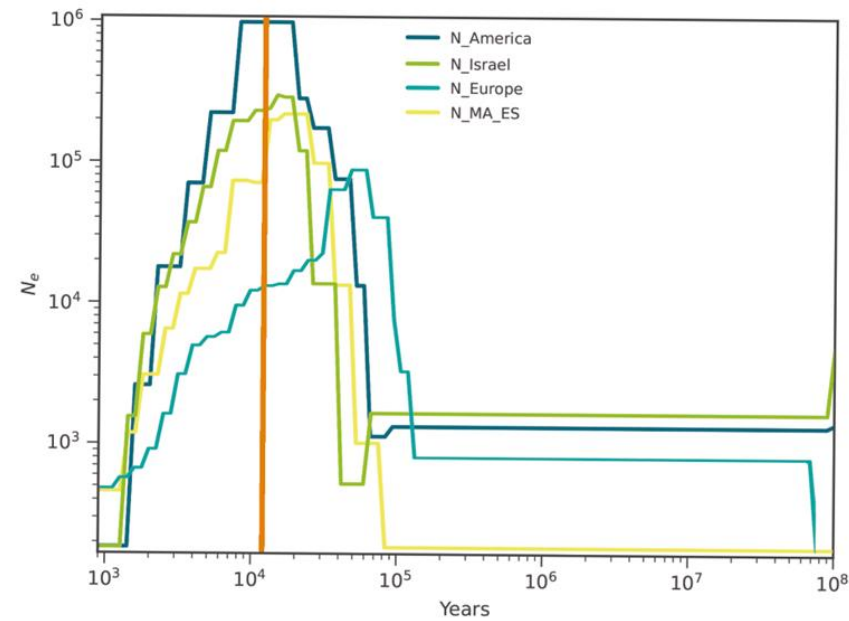


Figure S15. Demographic changes of N_e for four *D. plantaginea* genetic groups in the recent 10^8 years using SMC++ (Terhorst et al., 2017). The orange line corresponds to the Holocene period. N_{America} = *D. plantaginea* from Canada and the US; N_{Israel} = *D. plantaginea* from Israel; N_{Europe} = *D. plantaginea* from Belgium, Denmark, France, Germany, Romania, Spain, Sweden, and the UK; $N_{\text{MA_ES}}$ = *D. plantaginea* from Morocco and Spain.

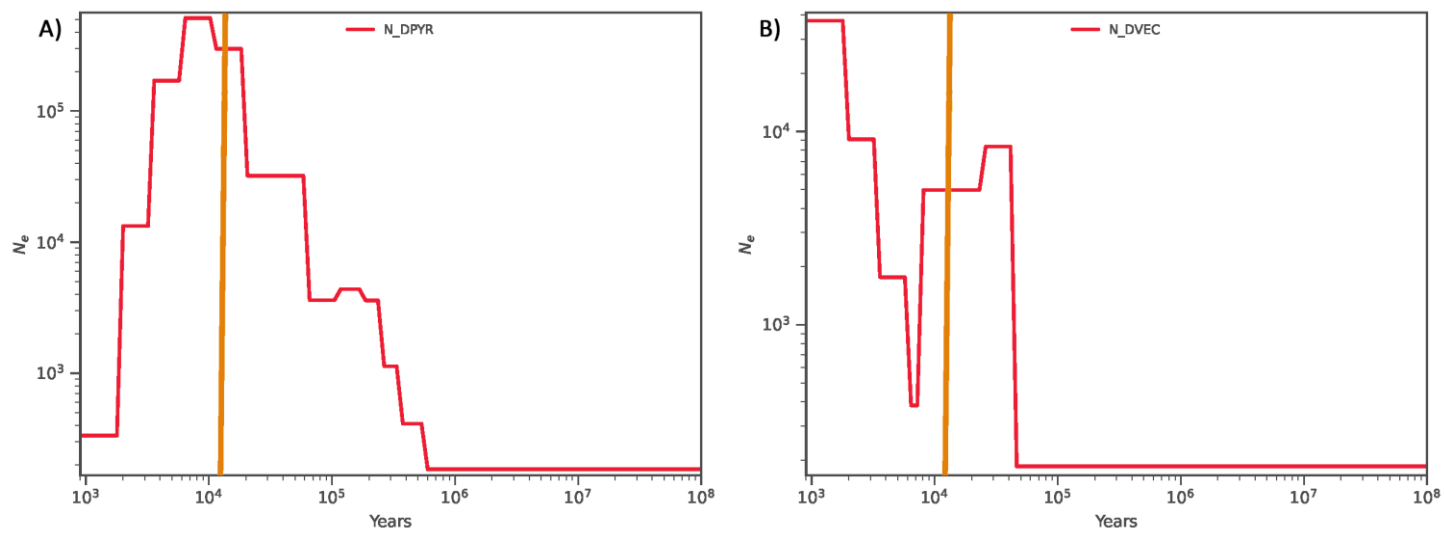


Figure S16. Demographic changes of N_e for A) *D. pyri* (N_DPYR) and B) *D. depecta* (N_DVEC) genetic groups in the recent 10^8 years using SMC++ (Terhorst et al., 2017). The orange line corresponds to the Holocene period.

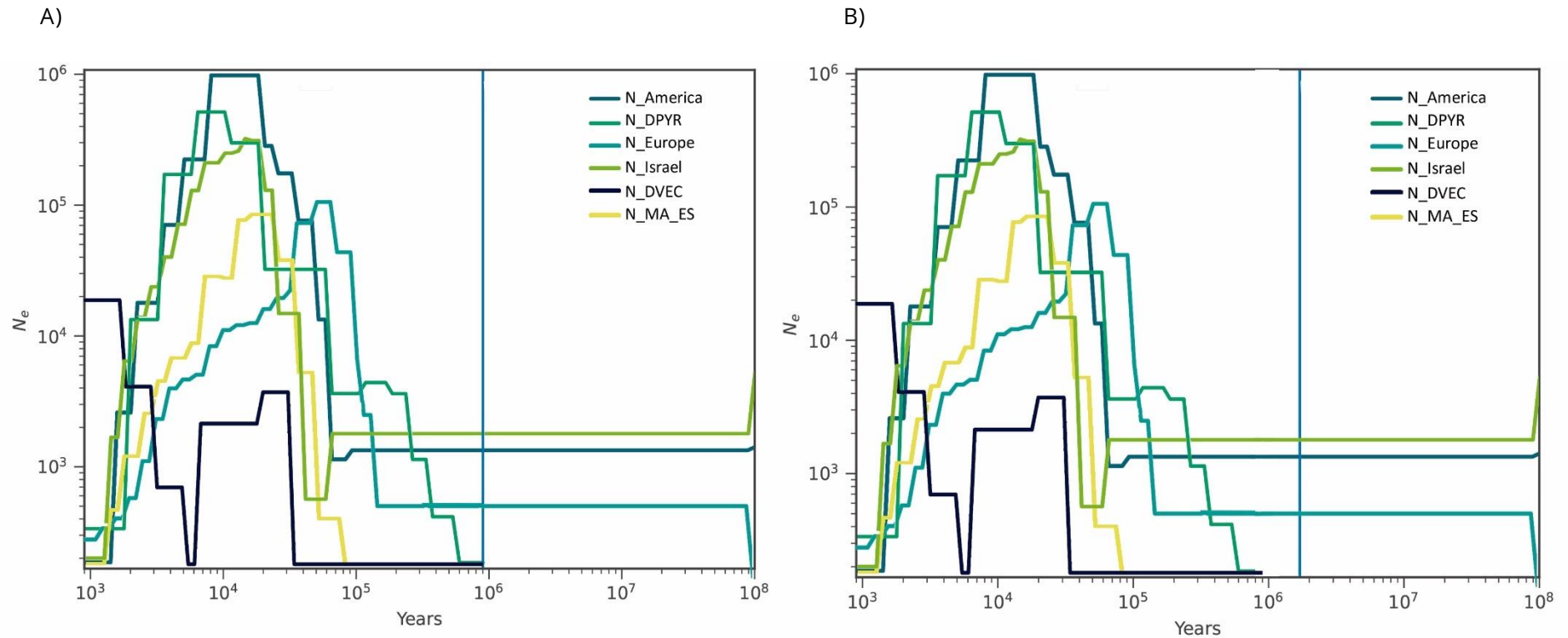


Figure S17. Splitting time of the aphid population genetic groups inferred for A) *D. pyri* in less than 10^6 years and B) *D. devecta* in later than 10^6 years using SMC++ (Terhorst et al., 2017). The blue line corresponds to the split time. The aphid genetic groups are N_America = *D. plantaginea* from Canada and the US; N_Israel = *D. plantaginea* from Israel; N_Europe = *D. plantaginea* from Belgium, Denmark, France, Germany, Romania, Spain, Sweden, and the UK; N_MA_ES = *D. plantaginea* from Morocco and Spain; N_DPYP = *D. pyri*; N_DVEC = *D. devecta*.

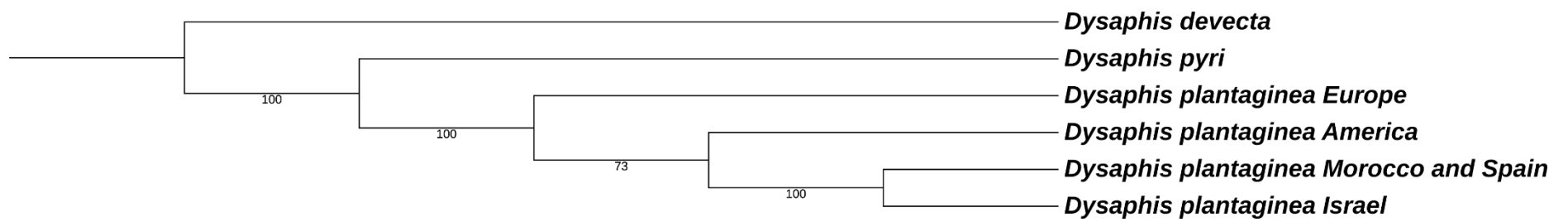


Figure S18. Aphid species-tree inference with the single value decomposition quartets (SVDQuartets) score (Chifman and Kubatko, 2014). Bootstrap values are described in the branches of the tree.

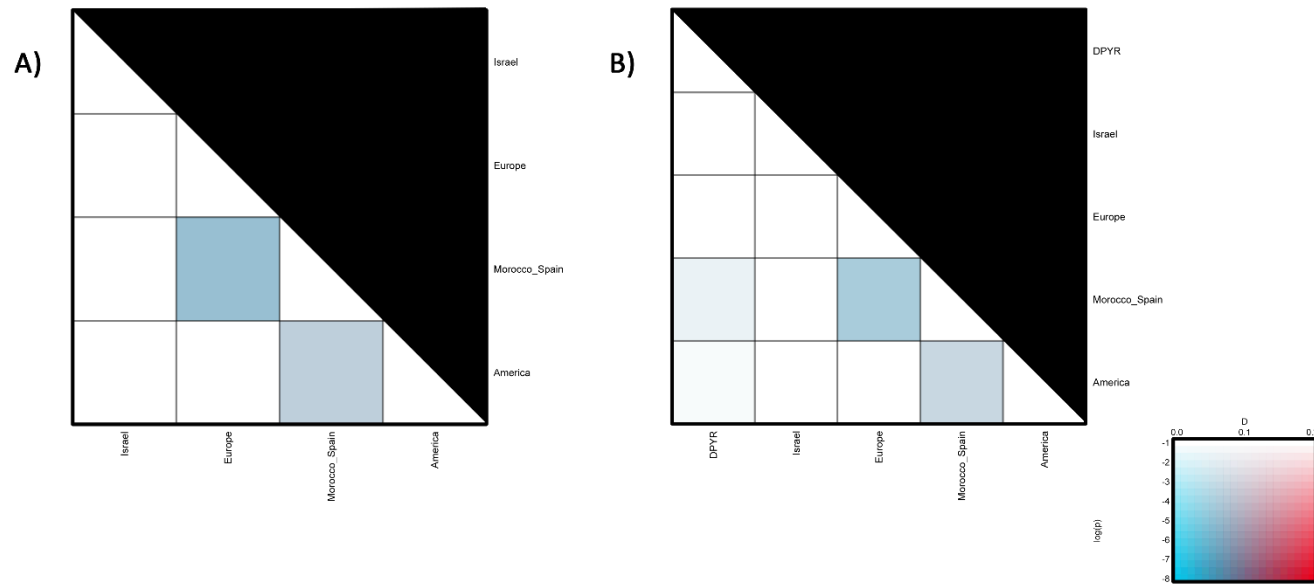


Figure S19. Heatmap of *D*-statistic values (Malinsky et al., 2020) of the tested hypotheses: A) Gene flow among *D. plantaginea* genetic groups. B) Gene flow between *D. plantaginea* and *D. pyri* genetic groups. The lower right panel shows the absolute *D*-statistic value in red color the higher while the lower in blue color.

Table S2. Summary of *D*-statistic of testing gene flow among *D. plantaginea* genetic groups (Hypothesis 1).

BBA												
P1	P2	P3	<i>D</i> statistic	Z-score	<i>p</i> -value	f4-ratio	ABBA	BABA	BBA	D_calculated	<i>q</i> _value	Stdv
Israel	America	Europe	0,003	0,178	0,859	0,000	244,587	243,009	252,244	0,003	0,859	0,162
America	Morocco_Spain	Europe	0,060	4,339	0,000	0,000	252,046	223,628	261,120	0,060	0,000	0,133
Israel	Morocco_Spain	America	0,071	3,666	0,000	0,403	225,725	195,699	250,932	0,071	0,001	0,205
Israel	Morocco_Spain	Europe	0,070	3,935	0,000	0,000	232,064	201,845	266,313	0,070	0,000	0,185
Dmin												
P1	P2	P3	<i>D</i> statistic	Z-score	<i>p</i> -value	f4-ratio	ABBA	BABA	BBA	D_calculated	<i>q</i> _value	
Israel	America	Europe	0,003	0,178	0,859	0,000	244,587	243,009	252,244	0,003	0,859	
Europe	America	Morocco_Spain	0,018	1,782	0,075	0,079	261,120	252,046	223,628	0,018	0,100	
America	Israel	Morocco_Spain	0,053	2,517	0,012	0,240	250,932	225,725	195,699	0,053	0,019	
Europe	Israel	Morocco_Spain	0,069	3,650	0,000	0,297	266,313	232,064	201,845	0,069	0,001	

Table S3. Summary of *D*-statistic of testing gene flow between *D. plantaginea* and *D. pyri* genetic groups (Hypothesis 2).

P1	P2	P3	<i>D</i> statistic	Z-score	<i>p</i> -value	f4-ratio	BBAA			D_calculated	<i>q</i> value	Stdv
							ABBA	BABA	BBAA			
Europe	America	DPYR	0,052	1,929	0,054	0,001	57,395	51,762	11383,40 0	0,052	0,107	0,456
Israel	America	DPYR	0,007	0,140	0,888	0,000	51,543	50,852	11382,80 0	0,007	0,891	0,541
America	Morocco_Spain	DPYR	0,005	0,136	0,891	0,000	50,468	50,008	11417,40 0	0,005	0,891	0,497
Israel	America	Europe	0,008	0,397	0,691	0,000	228,18 3	224,51 8	241,785	0,008	0,891	0,177
America	Morocco_Spain	Europe	0,057	3,956	0,000	0,000	235,57 5	210,32 9	251,655	0,057	0,001	0,138
Israel	Morocco_Spain	America	0,070	3,400	0,001	0,391	212,81 3	184,89 8	234,764	0,070	0,003	0,216
Europe	Israel	DPYR	0,046	1,152	0,249	0,001	55,713	50,771	11371,30 0	0,046	0,416	0,552
Europe	Morocco_Spain	DPYR	0,059	2,343	0,019	0,001	54,713	48,619	11405,70 0	0,059	0,055	0,522
Israel	Morocco_Spain	DPYR	0,012	0,269	0,788	0,000	48,028	46,876	11427,10 0	0,012	0,891	0,527
Israel	Morocco_Spain	Europe	0,072	3,790	0,000	0,000	216,78 2	187,64 8	254,781	0,072	0,001	0,196

DPYR = *D. pyri*

Table S3 continuation. Summary of *D*-statistic of testing gene flow between *D. plantaginea* and *D. pyri* genetic groups (Hypothesis 2).

P1	P2	P3	<i>D</i> statistic	Z-score	<i>p</i> -value	f4-ratio	Dmin			D_calculated	<i>q</i> value
							ABBA	BABA	BBA		
Europe	America	DPYR	0,052	1,929	0,054	0,001	57,395	51,762	11383,40 0	0,052	0,107
Israel	America	DPYR	0,007	0,140	0,888	0,000	51,543	50,852	11382,80 0	0,007	0,891
America	Morocco_Spain	DPYR	0,005	0,136	0,891	0,000	50,468	50,008	11417,40 0	0,005	0,891
Israel	America	Europe	0,008	0,397	0,691	0,000	228,18 3	224,51 8	241,785	0,008	0,891
Europe	America	Morocco_Spain	0,033	3,219	0,001	0,140	251,65 5	235,57 5	210,329	0,033	0,005
America	Israel	Morocco_Spain	0,049	2,149	0,032	0,222	234,76 4	212,81 3	184,898	0,049	0,079
Europe	Israel	DPYR	0,046	1,152	0,249	0,001	55,713	50,771	11371,30 0	0,046	0,416
Europe	Morocco_Spain	DPYR	0,059	2,343	0,019	0,001	54,713	48,619	11405,70 0	0,059	0,055
Israel	Morocco_Spain	DPYR	0,012	0,269	0,788	0,000	48,028	46,876	11427,10 0	0,012	0,891
Israel	Morocco_Spain	Europe	0,072	3,790	0,000	0,000	216,78 2	187,64 8	254,781	0,072	0,001

DPYR = *D. pyri*

REFERENCES

- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1). <https://doi.org/10.1186/s13742-015-0047-8>
- Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23), 3317-3324. <https://doi.org/10.1093/bioinformatics/btu530>
- Huson, D. H., & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23(2), 254-267. <https://doi.org/10.1093/molbev/msj030>
- Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic acids research*, 49(W1), W293-W296. <https://doi.org/10.1093/nar/gkab301>
- Li, Y. L., & Liu, J. X. (2017). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, 18(1), 176-177. <https://doi.org/10.1111/1755-0998.12719>
- Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Molecular ecology resources*, 21(2), 584-595. <https://doi.org/10.1111/1755-0998.13265>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37(5), 1530-1534. <https://doi.org/10.1093/molbev/msaa015>
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573-589. <https://doi.org/10.1534/genetics.114.164350>

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature genetics*, 49(2), 303-309. <https://doi.org/10.1038/ng.3748>

Wickham, H., François, R., Henry, L., Müller, K. (2022). *dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., & Yang, T. L. (2018). PopLDdecay : a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, 35(10), 1786-1788. <https://doi.org/10.1093/bioinformatics/bty875>

7. OUTLINE CHAPTER 4 : AN EXPERIMENTAL TEST FOR LOCAL ADAPTATION OF THE ROSY APPLE APHID (*DYSAPHIS PLANTAGINEA*) DURING ITS RECENT RAPID COLONIZATION ON ITS CULTIVATED APPLE HOST (*MALUS DOMESTICA*) IN EUROPE

Plant-parasite interactions have been seen as a major driver of local adaptation and, therefore, used as a model for studying the evolutionary processes involved in local adaptation (Garrido et al., 2012). Theoretical work predicts that variations in selection between environments can lead to patterns of local adaptation of the parasite (Kawecki and Ebert, 2004; Olazcagua et al., 2022). Parasite population is locally adapted when the sympatric populations (e.g., parasite infested on local host) have higher survival or performance (fitness) compared with allopatric populations (e.g., parasite infested on foreign host; Kaltz and Shykoff, 1998). A handful of experiments tested these theoretical expectations. In particular, common gardens are effective experimental designs to identify key traits underlying local adaptation and the ecological factors shaping the adaptation (Kaltz and Shykoff, 1998; Tiffin and Ross-Ibarra, 2014; Greischar and Koskella, 2007; Nuismar and Gandon, 2008; Nooten and Hughes, 2017). Cross-infection experiments with the fungal pathogens *Colletotrichum lindemuthianum* (Sacc. & Magnus) Briosi & Cavara (Sicard et al., 2007) and *Microbotryum violaceum* (Persoon) Deml & Oberwinkler (Kaltz and Shykoff, 1998; Kawecki and Ebert, 2004; Greischar and Koskella, 2007) revealed various patterns, from local adaptation (Kaltz et al., 1999) to maladaptation to their hosts. However, the contributions of the abiotic and biotic (i.e., plant host) environment to parasite local adaptation are still unknown. Multi-population manipulative transplant and cross-infection experiments to test the effect of the host (i.e., biotic) and the abiotic environments reflecting a spatial heterogeneity (e.g., spatially varying climate) of the parasite environments, therefore, remain to be done (Gandon et al., 2008; Brown and Tellier, 2011). In addition, the

handful of experimental tests of local adaptation to plant hosts focused on phytopathogens (fungi mostly) as they are easier to manipulate for controlled infections than other major parasite groups, such as phytophagous insects. Cross-infection and transplant experiments are also technically and human-challenging, making them rarely settled. Therefore, there is a need to understand better the evolutionary process behind the local adaptation of phytophagous insects to both biotic and abiotic conditions, which is essential to developing strategies for pest control (Kawecki and Ebert, 2004; Tiffin and Ross-Ibarra, 2014).

During my Ph.D., using aphids as models, I investigated the extent of adaptation of phytophagous insects to their local hosts and environments. As I stated in my introduction, aphids represent an ideal model for studying host-parasite interactions because of their strong connection with their host (Simon et al., 2015). Their host is their biotic environment, food provider, and habitat throughout their life cycle (Lee and Lee, 2020). Their life cycle depends on abiotic factors such as photoperiod to promote hatching, winged induction, and sexual aphid forms (Braedle et al., 2006). Other abiotic factors (e.g., temperature, light) determined asexual dominance or exclusivity in the aphid life cycle (Simon et al., 2002; Simon et al., 2010). However, more studies are needed about the ecological bases of the aphid local adaptation to their abiotic (*i.e.*, local environment) and biotic (*i.e.*, plant host).

In Chapter 4, we explore the responses of different cultivated apple tree genotypes to the infestation of local and foreign rosy apple aphid (*Dysaphis plantaginea* Passerini) genotypes from different origins. The experiment initially consisted of three common gardens representing an environment gradient: Belgium (mean annual temperature of 9.6°C and annual precipitation of 823 mm), France (mean annual temperature of 11.4°C and annual precipitation of 675 mm

annual precipitation), and Spain (mean annual temperature of 11.8°C and annual precipitation of 869 mm). We performed an aphid cross-infestation experiment from early spring to early summer 2021. We utilized 28 apple tree genotypes comprising 15 cultivated apple genotypes (five local genotypes from each country), nine wild apple genotypes (*Malus sylvestris* L. Mill., six from Belgium and three from Spain), and four apple genotypes with different susceptibility levels against aphid infestations: three tolerant apple genotypes (one 'Priscilla', one 'Florina', and one genotype of the ornamental species *Malus floribunda* Siebold ex Van Houtte), and one susceptible genotype, the *M. domestica* Golden Delicious. We sampled nine genotypes representing three rosy apple aphid populations from a country and exchanged them among the three common gardens during spring 2021. Note that the experiment was initially planned for 2020, but because of the COVID-19 lockdown, we postponed it to 2021, and I obtained a one-year extension for my Ph.D. thesis. We, therefore, performed the aphid infestation during the early summer of 2021. I was in Spain and led the experiment there, while other teams were on the two other sites I interacted with daily to synchronize the experiments. We placed three adult aphids on each apple genotype on nine different infestation spots (represented by a randomly selected leaf). We protected them with an aphid cage developed by the ECLECTIC team. After around 5 to 7 days, we recovered the aphid colonies and took a picture of each colony at the laboratory. The experimental design was published in Peer Community in Ecology to improve our experiment based on the suggestions of researchers with high expertise in the area (https://forgemia.inra.fr/amandine.cornille/local_adaptation_dp).

We had the opportunity to collaborate with the physics laboratory of Angers University, led by Dr. David Rousseau, and exchanged with the research engineer Herearii Metuarea to develop machine-learning tools to classify each aphid stage from the picture of the colony we sampled on the field. We defined a classification

based on the aphid morphology to detect the colony growth, the number of instars, and the surface of each individual in the colony.

Chapter 4 is in the process of writing and analysis as it took time to obtain the counting and measures for each colony from the machine learning tools developed with our colleagues from Angers University. Note that the experiment in Belgium is not included, as a big storm destroyed the infestation spots on this site in 2021. Therefore, we focused on the Spanish and French sites. I am presenting preliminary analyses to fulfill the deadline for the thesis submission. However, I will process the data to prepare a manuscript to publish this work in a journal.

OUTLINE REFERENCES

Braendle, C., Davis, G. K., Brisson, J. A., & Stern, D. L. (2006). Wing dimorphism in aphids. *Heredity*, 97(3), 192-199. <https://doi.org/10.1038/sj.hdy.6800863>

Brown, J. K., & Tellier, A. (2011). Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual review of phytopathology*, 49, 345-367. <https://doi.org/10.1146/annurev-phyto-072910-095301>

Gandon, S., Buckling, A., Decaestecker, E., & Day, T. (2008). Host-parasite coevolution and patterns of adaptation across time and space. *Journal of evolutionary biology*, 21(6), 1861-1866. <https://doi.org/10.1111/j.1420-9101.2008.01598.x>

Garrido, E., Andraca-Gómez, G., & Fornoni, J. (2012). Local adaptation: simultaneously considering herbivores and their host plants. *New Phytologist*, 193(2), 445-453. <https://doi.org/10.1111/j.1469-8137.2011.03923.x>

Greischar, M. A., & Koskella, B. (2007). A synthesis of experimental work on parasite local adaptation. *Ecology letters*, 10(5), 418-434. <https://doi.org/10.1111/j.1461-0248.2007.01028.x>

Kaltz, O., & Shykoff, J. A. (1998). Local adaptation in host-parasite systems. *Heredity*, 81(4), 361-370. <https://doi.org/10.1046/j.1365-2540.1998.00435.x>

Kaltz, O., Gandon, S., Michalakis, Y., & Shykoff, J. A. (1999). Local maladaptation in the anther-smut fungus *Microbotryum violaceum* to its host plant *Silene latifolia*: evidence from a cross-inoculation experiment. *Evolution*, 53(2), 395-407. <https://doi.org/10.1111/j.1558-5646.1999.tb03775.x>

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology letters*, 7(12), 1225-1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>

Lee, S., & Lee, S. (2020). Multigene phylogeny uncovers oviposition-related evolutionary history of Cerambycinae (Coleoptera: Cerambycidae). *Molecular phylogenetics and evolution*, 145, 106707. <https://doi.org/10.1016/j.ympev.2019.106707>

Nooten, S. S., & Hughes, L. (2017). The power of the transplant: direct assessment of climate change impacts. *Climatic Change*, 144, 237-255. <https://doi.org/10.1007/s10584-017-2037-6>

Nuismer, S. L., & Gandon, S. (2008). Moving beyond common-garden and transplant designs: insight into the causes of local adaptation in species interactions. *The American Naturalist*, 171(5), 658-668. <https://doi.org/10.1086/587077>

Olazcuaga, L., Foucaud, J., Deschamps, C., Loiseau, A., Claret, J. L., Vedovato, R., ... & Estoup, A. (2022). Rapid and transient evolution of local adaptation to seasonal host fruits in an invasive pest fly. *Evolution Letters*, 6(6), 490-505. <https://doi.org/10.1002/evl3.304>

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11), 807-820. <https://doi.org/10.1038/nrg3522>

Sicard, D., Pennings, P. S., Grandclément, C., Acosta, J., Kaltz, O., & Shykoff, J. A. (2007). Specialization and local adaptation of a fungal parasite on two host plant species as revealed by two fitness traits. *Evolution*, 61(1), 27-41. <https://doi.org/10.1111/j.1558-5646.2007.00003.x>

Simon, J. C., Rispe, C., & Sunnucks, P. (2002). Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, 17(1), 34-39. [https://doi.org/10.1016/S0169-5347\(01\)02331-X](https://doi.org/10.1016/S0169-5347(01)02331-X)

Simon, J. C., Stoeckel, S., & Tagu, D. (2010). Evolutionary and functional insights into reproductive strategies of aphids. *Comptes Rendus Biologies*, 333(6-7), 488-496. <https://doi.org/10.1016/j.crvl.2010.03.003>

Simon, J. C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquier, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in functional genomics*, 14(6), 413-423. <https://doi.org/10.1093/bfgp/elv015>

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12), 673-680. <https://doi.org/10.1016/j.tree.2014.10.004>

7.1 AN EXPERIMENTAL TEST FOR LOCAL ADAPTATION OF THE ROSY APPLE APHID (*DYSAPHIS PLANTAGINEA*) DURING ITS RECENT RAPID COLONIZATION ON ITS CULTIVATED APPLE HOST (*MALUS DOMESTICA*) IN EUROPE (MANUSCRIPT IN PREPARATION)

Olvera-Vazquez S.G.¹, Alhmedi A.², Miñarro M.³, Shykoff J. A.⁴, Marchadier E.¹, Rousselet A. ¹, Remoué C.¹, Gardet R.⁵, Degrave A. ⁵, Robert P. ⁵, Boulez A. ¹, Chen X.¹, Porcher J. ⁵, Vander-Mijnsbrugge K.⁶, Raffoux X. ¹, Falque M. ¹, Anaya-Sainz J.M.¹, Gay R.¹, Alins, G.⁷, Didelot F.⁸, Beliën T.², Dapena E.³, Lemarquand A. ⁸, Cornille A.¹

¹ Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190 Gif-sur-Yvette, France

² Department of Zoology, pcfruit vzw, Sint-Truiden, Belgium

³ Servicio Regional de Investigación y Desarrollo Agroalimentario (SERIDA), Carretera AS-267, PK. 19, E-33300, Villaviciosa, Asturias, Spain

⁴ Laboratoire d'Ecologie, Systématique et Evolution, Université Paris-Saclay, CNRS, AgroParisTech, 91400 Orsay cedex, France.

⁵ AgroCampus-Ouest, UMR1345 Institut de Recherche en Horticulture et Semences (IRHS), 49045 Angers, France

⁶ Department of Forest Ecology and Management, Research Institute for Nature and Forest, 9500 Geraardsbergen, Belgium

⁷ IRTA Fruitcentre, PCiTAL, Park of Gardeny, Fruitcentre Building, 25003 Lleida, Spain

⁸ INRAE, Unité Expérimentale Horticole N34 0449, Centre d'Angers-Nantes, 49071 Beaucouzé Cedex, France

ABSTRACT

Plant-parasite interactions have been seen as a major driver of local adaptation and, therefore, used as a model for studying the evolutionary processes involved in local adaptation. The relative contributions of biotic and abiotic factors to parasite local adaptation still need to be determined. Here, we investigated whether local adaptation occurred during the recent rapid colonization of cultivated apples by *Dysaphis plantaginea*, the major aphid pest of cultivated apple orchards in Europe. We experimental test for local adaptation of *D. plantaginea* by quantifying the fitness of three aphid populations from Belgium, France, and Spain infested in three common garden orchards in Belgium, France, and Spain. Each common garden comprised a panel of cultivated apple varieties from Belgium, France, and Spain. The results of the present work will generate original results to understand how the biotic (the host) and abiotic conditions can shape local adaptation in crop parasites. Understanding local parasite adaptation processes is essential for guiding pest management and breeding programs in the context of global changes.

Keywords: local adaptation, aphid, fruit trees, common garden, G*G*E interaction, host-parasite interaction, domestication

7.2 INTRODUCTION

The mode of action of natural selection in shaping the genetic diversity of the individuals of a population to allow them to adapt to their local environment is a central question in evolutionary biology. Populations are locally adapted to their environment when their fitness is higher in their local environment than elsewhere (Mopper and Strauss, 2013). However, it is also possible to observe maladaptation patterns when local populations show lower fitness in their local environment compared with foreign environments (Galloway and Fenster, 2000; Herford and Winn, 2008). Common gardens experiments have been a tool to detect key traits underlying local adaptation and the ecological factors shaping the adaptation, mainly in plants to their abiotic environment (Kaltz and Shykoff, 1998; Tiffin and Ross-Ibarra, 2014; Greischar and Koskella, 2007; Nuismar and Gandon, 2008; Nooten and Hughes, 2017). However, a few studies also focused on parasites.

Plant-parasite interactions have been seen as a major driver of local adaptation and are an important model for studying the evolutionary processes involved in local adaptation (Garrido et al., 2012). Theoretical work predicts that spatiotemporal fluctuations in selection pressures would lead to patterns of local adaptation of the parasite to its host (Kawecki and Ebert, 2004; Olazcagua et al., 2022). Evidence of local adaptation to the host has been detected using cross-infection experiments in plant pathogens. Patterns of local adaptation of *Colletotrichum lindemuthianum* (Sacc. & Magnus) Briosi & Cavara to its host *Phaseolus vulgaris* L. (Sicard et al., 2007) have been observed using fungi infectivity and aggressivity as fitness estimates. Patterns of local adaptation but also maladaptation to their host were identified in the plant fungal model system *Microbotryum violaceum* (Persoon) Deml & Oberwinkler (Kaltz and Shykoff, 1998; Kaltz et al., 1999; Kawecki and Ebert, 2004; Greischar and Koskella, 2007; Koupilová et al., 2021).

Local maladaptation was detected at the host species level in the phytophagous insect parasite, the soft-scale insect *Saissetia coffeae* Walker, using fecundity as a measure of fitness (Hanks et al., 1994). Populations of *S. coffeae* fit better on the original host, *Witheringia* L'Her spp, than on their novel host, *Lomariopsis* Fée spp (Spitzer et al., 2006). All those studies tested parasites local adaptation to their host but not to their biotic environment. But so far, the contributions of the abiotic (*i.e.*, local environment) and biotic (*i.e.*, local plant host) environments to parasite local adaptation are still unknown. Understanding the evolutionary process behind the local adaptation of parasites to both biotic and abiotic conditions is essential to developing strategies for pest control (Kawecki and Ebert, 2004; Tiffin and Ross-Ibarra, 2014).

Aphids are phytophagous parasites ideal for studying host-parasite interactions (Simon et al., 2015). The aphid host represents their biotic environment, the food source, and where their life cycle is developed (Lee and Lee, 2020). The abiotic factors are essential for transition among the different instars in the aphid life cycle; for instance, photoperiod is key for the hatching, winged induction, and sexual aphid forms (Braedle et al., 2006). Temperature and light are essential determinants of asexual predominance or exclusivity in the aphid life cycle (Simon et al., 2002; Simon et al., 2010). Nevertheless, the ecological bases of the aphid local adaptation to their abiotic (*i.e.*, local environment) and biotic (*i.e.*, plant host) must be clarified.

The rosy apple aphid or *Dysaphis plantaginea* Passerini is a major pest of the cultivated apple, causing up to 30% of annual economic losses, causing crinkled yellowish pseudogalls on leaves, small and malformed fruits (Ismail et al., 2022). Population genetics evidence (Olvera-Vazquez et al. 2021) showed quick

and recent colonization of the rosy apple aphid on its cultivated apple host in Europe when this fruit tree was brought there from Central Asia by the Romans and Greeks about 1,500 years ago (Cornille et al. 2014, 2019). Previous studies suggested that the apple genotypes tolerant to the rosy apple aphid infestations (*Malus floribunda* Siebold ex Van Houtte, *Malus domestica* Borkh Florina, and *M. domestica* Priscila) induced lower fitness of the rosy apple aphid compared to susceptible genotype *M. domestica* Golden Delicious (Miñarro and Dapena, 2007; Pagliarani et al., 2016; Dall'Agata et al., 2018).

Our question was, is there a pattern of local adaptation of the rosy apple aphid (*D. plantaginea*) to its local environment and/or cultivated apple host (*M. domestica*)? For this aim, we defined the following questions: do the rosy apple aphid genotypes from three different origins (Belgium, France, and Spain) show higher or lower fitness in their local environment (*i.e.*, Belgium, France, and Spain, respectively) compared to the fitness in their foreign environment? Do the rosy apple aphid genotypes from three different origins (*i.e.*, Belgium, France, and Spain) show higher or lower fitness on their respective local apple host genotypes (*i.e.*, local Belgian, French, and Spanish apple genotypes, respectively) compared to the fitness on their foreign apple genotypes?

7.3 MATERIAL AND METHODS

The present experiment was submitted for review by experts in the ecology of evolution subject for its improvement before its execution. The experimental design publication Olvera-Vazquez et al. (2021) describes more details of the present experiment.

COMMON GARDENS

The experiment was located at three common garden orchards at 1) Sint-Truiden in Belgium (50°48'0" N, 5° 11'0" E), presenting a mean annual temperature of 9.6°C and annual precipitation of 823 mm, 2) Beaucouzé in France (47°28'57" N, 0°36'52" W), presenting a mean annual temperature of 11.4°C and annual precipitation of 675 mm annual precipitation, and 3) Villaviciosa in Asturias in Spain (43°28'45" N, 5° 26'32" W), presenting a mean annual temperature of 11.8°C and annual precipitation of 869 mm. The bioclimatic information was extracted from the WorldClim - Global Climate database (<https://www.worldclim.org/>, Fick et al., 2017) with the raster R package (Hijmans and van Etter, 2012). Each common garden orchard contained 10 to 12 clones of 28 apple genotypes (described below). These were planted in 10 to 12 rows, each comprising the available genotypes placed randomly.

APHID SAMPLES

We sampled nine aphid genotypes (three from each country, *i.e.*, Belgium, France, and Spain) on locally cultivated apple trees in each experimental field during spring 2021. We got clonal offspring of a single female (called hereafter matriline) with three matriline from each country. We kept the aphids in a greenhouse infesting apple Jonagold and Golden delicious genotypes provided by the CRA-W company (Micropropagation laboratory, Biological Engineering Unit, Gembloux, Belgium). We kept a single aphid genotype per Jonagold or Golden Delicious tree under the following conditions: 18°-20°C, 60-65% relative humidity, and natural light conditions (about 16 hours of light and 8 hours of dark). The apple genotype used for aphid rearing was chosen to be different from any cultivars in the infestation experiment of the Spring of 2021 to avoid any aphid acclimatization to a specific apple genotype.

In March 2021, we sent to each local laboratory in Belgium, France, and Spain some progeny of each of the nine-matriline rosy apple aphid colonies. Locally, each lab was rearing and synchronizing the nine colonies in a greenhouse onto Jonagold and Golden Delicious genotypes for the infestation experiment. For the synchronization, we placed each of the nine aphid genotypes on Golden Delicious trees grafted onto an M9 Pajam® under the same conditions as the rearing. After six days of colony growth, we started the aphid synchronization for each aphid genotype using a single adult aphid. The aphid synchronization aimed to ensure the same developmental stage of the females and larvae infested on a plant. We were recovering the synchronized aphids every three to four days.

APPLE TREES

Each common garden comprised 28 apple genotypes (*Malus sp* Mill., Figure 1, Table S1). The selection of the apple genotypes is described in the supplementary material text 1. These apple genotypes comprised 15 cultivated apple genotypes (five local genotypes from each country), nine wild apple genotypes (*Malus sylvestris* L. Mill., six from Belgium and three from Spain), and four apple genotypes with different susceptibility levels against aphid infestations: three tolerant apple genotypes (one 'Priscilla', one 'Florina', and one genotype of the ornamental species *Malus floribunda* Siebold ex Van Houtte), and one susceptible genotype, the *M. domestica* Golden Delicious. We grafted 10 to 12 clonemates for each of the 28 apple genotypes in 2018 (Figure 1, Table S1). We also grafted 206 clonemates of the Golden Delicious genotype to get at least 60 trees per locality available for aphid rearing. In total, 1,157 apple trees were grafted in early 2019 on an M9 Pajam 2® apple rootstock and maintained for one year (February 2019-2020) at an outdoor nursery at La Retuzière, Beaucouzé, Angers, France (47°28'57" N, 0°36'52"

W). In early February 2020, the trees were transferred and planted in the three common garden orchards (Figure 1). Each tree was sprayed with Teppeki® (flonicamida 50%) insecticide, a Bordeaux mixture (20% copper) fungicide, DELFIN® (*Bacillus thuringiensis* sp. *Kurstaki*) anti-lepidopterous, Essen'ciel (orange essential oil) insecticide and fungicide, Karate Zeon® (Lambda cihalotrin 1.5%) and Movento® (Spirotetramat 15% p/v OD) insecticides, and Sokalcarbion WP® (calcined kaolin), a physical mineral barrier between pest and plants. These treatments were applied until the beginning of the experiment (March 2021).

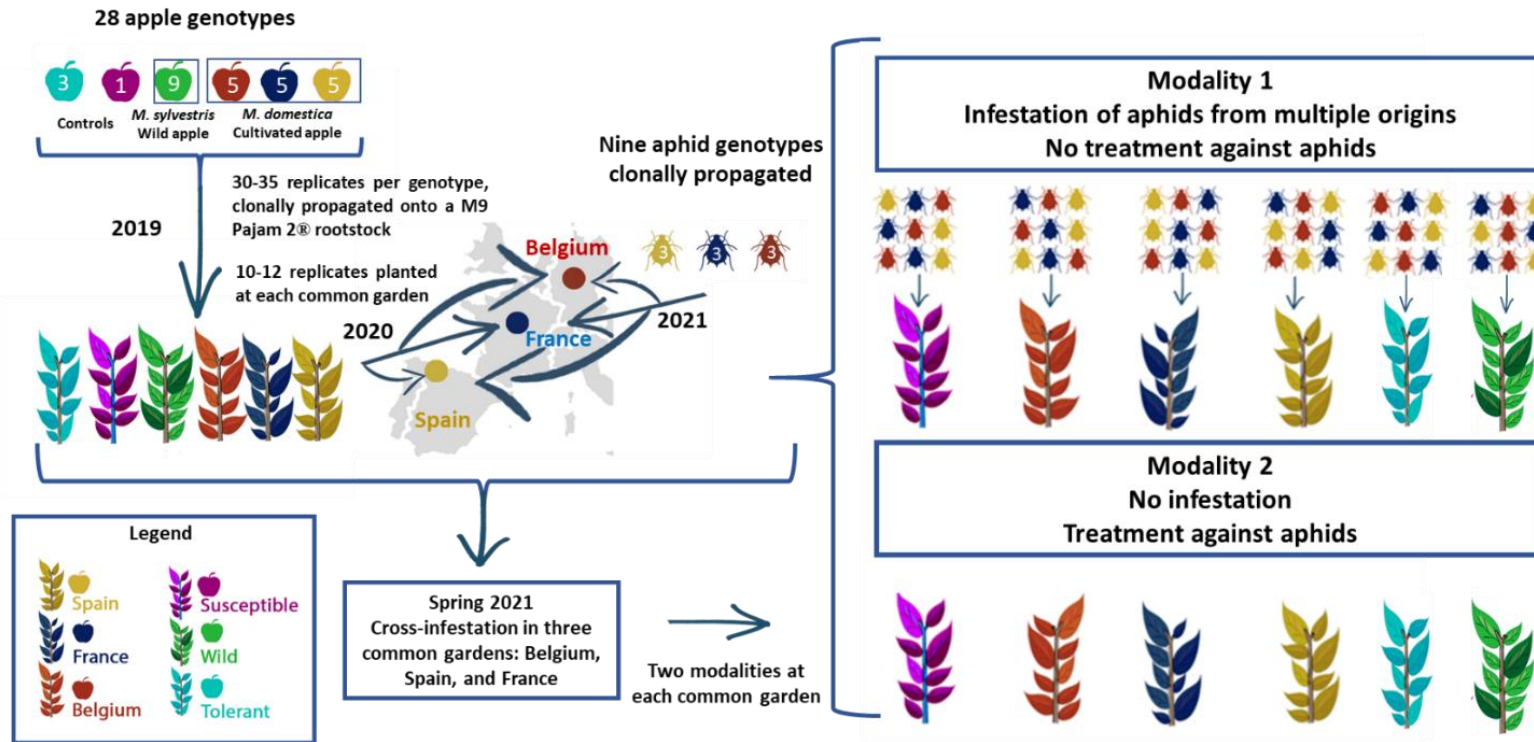


Figure 1. General scheme of the aphid cross-infestation experiment performed in the Spring of 2021 at the three common garden orchards in Belgium, Spain, and France. At each common garden orchard, 28 clonally propagated apple genotypes were grown with 10 to 12 replicates per genotype, depending on the survival of the grafted trees at each common garden. The apple genotypes included 1) *Malus domestica* genotypes from Belgium (five genotypes, red color), France (five genotypes, dark blue color), and Spain (five genotypes, yellow color). Additionally, 2) nine wild apple genotypes (*Malus sylvestris*), including six from Belgium and three from Spain (light green color), 3) *M. domestica* genotypes (Priscilla and Florina cultivars), and *Malus floribunda* Siebold ex Van Houtte, used as “tolerant to aphid infestation” controls (light blue color), and 4) the Golden delicious *M. domestica* genotype was used for aphid rearing as well as the “susceptible to aphid infestation” control (purple). Meanwhile, nine rosy apple aphid genotypes (*Dysaphis plantaginea*) were collected and clonally propagated: three from Belgium (red color), three from France (dark blue color), and three from Spain (yellow color). A total of 10-12 replicates of each of the 28 apple genotypes were transferred in February 2020 to each of the three common gardens. The aphid genotypes were transferred for rearing locally in February 2021 at each site.

GLOBAL DESIGN AND SAMPLING SIZE

We performed the aphid cross-infestation experiment in the spring of 2021. The planted apple genotypes were two years old since their transplantation, having acclimatized to their field conditions in the common garden for one year. We infested an apple leaf with two adult aphids and two larvae (aphids from the same colony previously reared and synchronized as we described in the aphid sample section). Each of the nine rosy apple aphid genotypes was placed on a different leaf on the same apple tree of each of the 28 apple genotypes in the three common garden orchards. The infestation was performed at the apple phenological stage E2 when the development of the inflorescences occurs (Bloesch et al., 2013). Aphid genotypes were placed on the leaves randomly for each level of the tree (upper, middle, lower). The infested leaf was protected with an aphid clip cage we developed for the experiment (Text S1). We performed the infestation in 18 days in each common garden. We recorded the date of initiation of each infestation and included these in the analyses as temporal blocks and the time within the days as a covariate. We tested for local adaptation of the rosy apple aphid to the cultivated apple and its local environment confronting the aphid fitness against 1) five cultivated apple genotypes from its native range, 2) 10 cultivated apple genotypes from two different non-native ranges, 3) nine wild apple genotypes, and 4) three apple genotypes tolerant to rosy apple aphid infestations (two *M. domestica* and one *M. floribunda*). In addition, each aphid genotype experienced climatic conditions from its native origin and two different local environmental conditions (including abiotic and biotic factors, such as climate or soil composition, and attacks of local parasites, respectively).

The experiment was divided into two modalities: Modality 1 with aphid infestation and free of treatment against aphids. This modality comprised one apple

genotype infested by the rosy apple aphid genotypes from different origins and seven to nine replicates of the 28 genotypes. The infestation was represented by a mini colony (two females and two larvae) by each of the nine aphid genotypes on nine different leaves on each of the 28 apple genotypes. Each mini colony was isolated and protected using an aphid clip cage. Each leaf was infested with a single aphid genotype from either Belgium, France, or Spain. Because the aphid life cycle varied with the climatic conditions among sites at each location, we utilized the duration of the aphid life cycle from adult to daughter-adult on a “time infestation control” cultivated apple genotype, *i.e.*, a susceptible Golden Delicious genotype (Miñarro and Dapena, 2008). We started the Golden Delicious apple tree infestation as a “reference” to determine the site’s standard duration of an aphid infestation (time to wait after an infestation to collect the colonies for each location). As previously published, this duration was usually between seven to 12 days after the initial infestation (Warneys et al., 2018). After reaching the colony growth time, we cut off each infested leaf with the clip cage. Then, we disassembled the clip cage to take the leaves with the aphid colony and transferred them into a Falcon tube previously labeled and filled with isopropanol. The second modality (Modality 2 hereafter) consisted of apple genotypes free of rosy apple aphid infestations and three replicates of the 28 genotypes used as non-infested controls.

From the beginning of the experiment, we defined 216 sympatric combinations and 423 allopatric combinations, which provides adequate power for testing local adaptation (Kaltz and Shykoff, 1998; Kaltz et al., 1999): we had 2/3 of allopatric comparisons (*i.e.*, aphid genotypes infested on their foreign apple genotypes and environments) against 1/3 sympatric comparisons (*i.e.*, aphid genotypes infested on their local apple genotypes and environments). However, natural events during the experiment reduced the number of infestations (see results section).

MEASUREMENTS AND STATISTICS

APHID MEASUREMENTS

For each aphid colony of each of the nine rosy apple aphid genotypes infested on the 28 apple genotypes, we counted the different insect life stages (*i.e.*, aphid larvae (L1 to L5)), apterous adults, and winged forms (Angeli and Simoni, 2006) using a method based on Napari-ImageJ (Selzer et al., 2023) in collaboration with the physics laboratory of Angers University, led by Dr. David Rousseau and developed with the research engineer Herearii Metuarea (Text S2).

We then defined aphid fitness estimates (W) calculated as follows (Warneys et al., 2018):

$$W_{growthRate} = \frac{n(\text{nymphs at end of infestation}) - n_{aphid}(\text{nymphs at beginning of infestation})}{\text{total number of day of infestation}}$$

We used a generalized linear mixed model (GLMM) with different factors to test our scientific question as follows (Table S2 describes the indexes, terms, and effects of our proposed GLMM Where):

$$W_{hijklmnot2z} = \mu_W + \text{aphid_origin}_h + \text{apple_origin}_i + \text{site}_j + \text{site}_j(\text{block}_k) + G_{h_i}(\text{leaf}_m(Gp_n)) \\ + \text{day_of_infestation}_t + \text{hour_of_infestation}_{t2} + \text{leaf_level}_m + \text{tree_clone}_o + \\ \text{aphid_origin}_h * \text{site}_j + \text{aphid_origin}_h * \text{apple_origin}_i + \text{aphid_origin}_h * \text{apple_origin}_i * \text{site}_j + \epsilon_{hijklmnot2z}$$

(Model 1)

$W_{hijklmnot2z}$ is the absolute fitness value of an aphid genotype Gp (*i.e.*, parasite genotype) from the country of origin n infested on the apple genotype l , apple tree clone o , in block k on *leaf level* m and the common garden j infested at day t and

hour t_2 , $leaf_level_m$ is the position of the infested leaf in the apple tree (upper, middle or lower), $tree_clone_o$ is the clone o of the apple genotype l , μ_w is the mean absolute fitness, $site_j$ is the common garden location (Belgium, Spain, France), $block_k$ is the block effect within each site for modality 1, $aphid_origin_h$ is the country of origin of the aphid (Spain, France, Belgium), $apple_origin_i$ is the country of origin of the apple genotype (Spain, France, Belgium), Gh_l is the apple genotype (*i.e.*, apple cultivar name) and $\varepsilon_{hijklmnott2z}$ is the residual term. *Block* is random and nested within the site, *aphid genotype* _{n} is nested within *leaf_level* _{m} , and *leaf_level* _{m} is nested within apple genotype Gh_l , and they were added to the models as random-effect terms. The *leaf_level* _{m} effect is also counted as a random factor to account for the global variability in aphid fitness explained by the levels at which each aphid colony was infested, whatever the apple genotypes. The *site* term measures the quality or suitability of the common garden locations, *aphid_origin* and *apple_origin* account for differences in fitness intrinsic to each local aphid genotype and apple genotype country of origin, $aphid_origin_h * site_j$ accounts for differences in local adaptation to the environment among the three aphid origins, $aphid_origin_h * apple_origin_i$ account for differences in local adaptation to the host among the three aphid origins, $aphid_origin_h * apple_origin_i * site_j$ accounts for differences in local adaptation to the host and environment among the three aphid origins. The $day_of_infestation_t$ and the $hour_of_infestation_{t_2}$ consider the effect of the infestation time of the aphid genotype Gp from the country-of-origin n on the apple genotype l in block k on leaf m and in the common garden j . We run our proposed model using one absolute fitness (W) measure: colony growth rate. We gradually removed interactions and effects according to their significance. In addition, we evaluated the differences in the effect on aphid fitness using a contrast analysis.

7.4 PRELIMINARY RESULTS

From the initial targeted 6,408 infestation spots, we had 2,679 infestation spots (two adult apterous and two larvae on a single apple genotype leaf) in the three common gardens: 257 in Belgium, 1,036 trees in France, and 1,386 in Spain. This decrease in the number of infestations reflects the challenge of this ambitious experiment. During the aphid cross-infestation experiment from early spring to early summer 2021, we faced unpredictable extreme local weather conditions and the presence of other insect organisms such as earwigs, predators of the aphids, and other species of aphid species among the three common gardens that impacted our aphid clip cages. Cold spells, heat waves, storms, or floods were reported in 2021 because of the climatic change reported by the European Geosciences Union (Faranda et al., 2022) and the World Meteorological Organization (World Meteorological Organization, 2022). As a result of this situation, we had to reduce our data universe, eliminating the Belgian site and genotypes because of the low number of representative infestation spots in this common garden. This has been an interesting challenge for me to understand how to process ecological data.

Our preliminary results showed an interaction between the common garden location and the origin of aphids (France and Spain) on aphid fitness (Table 1).

Table 1. Summary of Model 1 (GLMM, Negative Binomial) to investigate variation in the growth rate of the rosie apple aphid colony in response to its host (country_of_origin_tree) and local environment (common garden). *P-values* are explained below in the Table (*: $P < 0.05$ **: $P < 0.001$). The variable response was the

colony growth rate. Origin_aphid: origin of the aphid genotype, French or Spanish; Common_garden: common garden, French or Spanish; Country_origin_tree: origin of the tree genotype, French or Spanish, Resistant or tolerant.

Response: Colony Growth Rate			
	X ²	D.F.	Pr>X ²
Origin_aphids	3.270	1	0.070
Common_garden	11.334	1	0.001***
Country_origin_tree	2.425	4	0.658
Origin_aphid: common_garden	4.409	1	0.036*

We further investigated in each common garden the differences in colony growth rates and showed that the latter interaction was mainly explained by differences in growth rates in the French site between the Spanish and French aphid populations (Figure 1, Tables S3 and S4). We also compared the fitness of each aphid population among common gardens (Tables S5 and S5) and showed differences in fitness of the French aphid population, showing higher fitness in its foreign common garden. Therefore, the rosy apple aphid genotypes from France showed lower fitness in their local environment compared to their foreign common garden, while the Spanish aphid did not show different fitness depending on the orchards. These preliminary results suggest maladaptation of the French aphid to the local common garden conditions while global adaptation of the Spanish population. However, we did not see any effect of the origin of the apple population on aphid fitness, suggesting no local adaptation of the rosy apple aphid to the cultivated apple host.

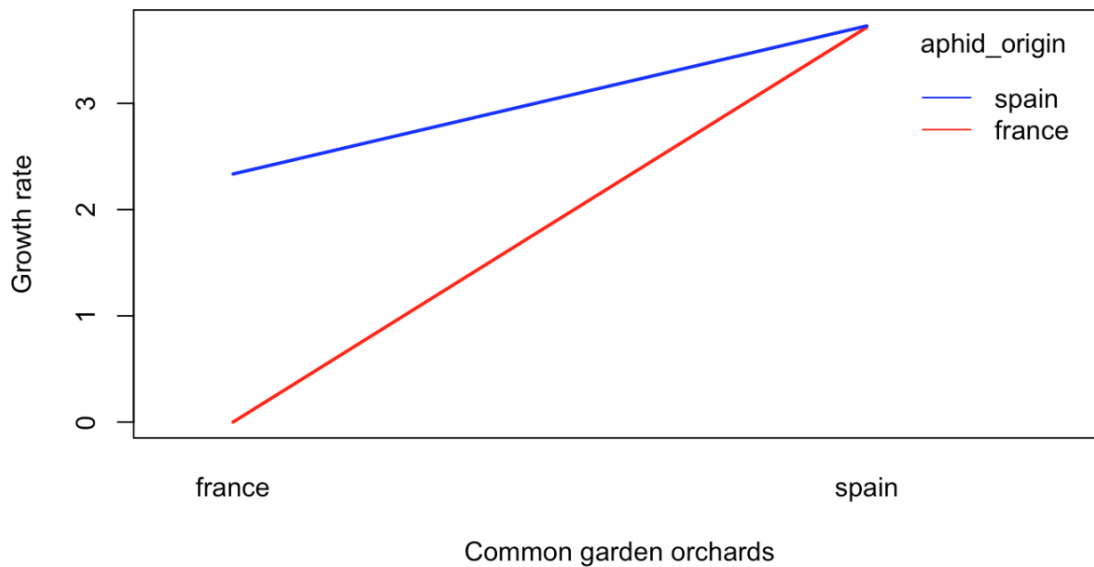


Figure 1. Interaction plot of growth rates of French and Spanish aphid colonies of the French (red color) and Spanish (blue color) common gardens.

7.5 DISCUSSION AND PERSPECTIVES

The experiment presented was a human and scientific adventure. From the settlement of the trees in the different countries to the infestation on the field and synchronization among sites and data processing afterward. For data processing, we wanted to avoid bias in aphid counting and size assessment, so we collaborated with the University of Angers. The development of the machine-learning tool implemented in Napari took almost one year. It allows us to get precise counts and size estimates for each aphid. Our preliminary results suggest different responses of aphid populations to their local sites but not hosts, but further investigations are needed. Indeed, our current results showed interactions between the common garden and the origin of the aphid genotypes, which are explained by the differences in the fitness of the French and Spanish aphid genotypes in the French site, which could suggest maladaptation patterns. We are considering

other fitness measurements to find patterns of local adaptation, such as the stress of the aphid colony estimated by the number of winged forms and precursors of winged forms and the size of the surface of each aphid in each colony. We are also processing the data obtained from the cultivated apple to understand the host response against the local environment and its aphid pest. The findings of the present chapter will complement the population genomic study of Chapter 3 to elucidate the role of abiotic and biotic factors involved in the adaptation of the rosy apple aphid.

ACKNOWLEDGMENTS

We are grateful to Charles Eric for his assistance in the apple genotype selection. We thank the equipment and technical support of the Fablab Digiscope-LISN-Université Paris-Saclay key in producing the aphid clip cages. We would also like to thank the team of institutions that are giving us invaluable support in organizing and performing the present experiment. The institutions are the GQE-Le Moulon, IRHS-INRAE-Agrocampus-Ouest-Université d'Angers in France, SERIDA in Spain, and PCFruit in Belgium. As well as interns involved in the experiment performance, Boulez Aurelien, Cristina de los Reyes, Ramón Franco, Etienne Porquier, Lawrence Pelgrims, Marie Bouchaud, Quentin Marcou, Valerio Cosemans.

FUNDINGS

The ATIP-Inserm-CNRS program supported this research. The University Paris-Saclay Short-Term Fellowship fund supported SGOV stay in Spain.

CONFLICT OF INTEREST DISCLOSURE

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

DATA, SCRIPTS, AND AVAILABILITY

Data are available online on: to be announced.

7.6 REFERENCES

- Angeli, G. I. N. O., & Simoni, S. (2006). Apple cultivars acceptance by *Dysaphis plantaginea* Passerini (Homoptera: Aphididae). *Journal of Pest Science*, 79(3), 175-179. <https://doi.org/10.1007/s10340-006-0129-6>
- Bloesch, B., & Viret, O. (2013). Stades phénologiques repères des fruits à pépins (pommier et poirier). *Revue suisse de viticulture, arboriculture et horticulture*, 45(2), 128-131. https://www.revuevitiarbohorti.ch/wp-content/uploads/2013_02_f_345.pdf
- Cornille, A., Giraud, T., Smulders, M. J., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, 30(2), 57-65. <https://doi.org/10.1016/j.tig.2013.10.002>
- Cornille, A., Antolín, F., Garcia, E., Vernesi, C., Fietta, A., Brinkkemper, O., ... & Roldán-Ruiz, I. (2019). A multifaceted overview of apple tree domestication. *Trends in plant science*, 24(8), 770-782. <https://doi.org/10.1016/j.tplants.2019.05.007>
- Dall'Agata, M., Pagliarani, G., Padmarasu, S., Troglio, M., Bianco, L., Dapena, E., ... & Tartarini, S. (2018). Identification of candidate genes at the Dp-fl locus conferring resistance against the rosy apple aphid *Dysaphis plantaginea*. *Tree genetics & genomes*, 14(1), 12. <https://doi.org/10.1007/s11295-018-1227-3>
- Faranda, D., Bourdin, S., Ginesta, M., Krouma, M., Noyelle, R., Pons, F., Yiou, P., and Messori, G.: A climate-change attribution retrospective of some impactful weather extremes of 2021, *Weather Clim. Dynam.*, 3, 1311–1340, <https://doi.org/10.5194/wcd-3-1311-2022>, 2022
- Fick, S. E. and Hijmans R.J., 2017. WorldClim 2: New 1-Km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12),4302-4315. <https://doi.org/10.1002/joc.5086>
- Galloway, L. F., & Fenster, C. B. (2000). Population differentiation in an annual legume: local adaptation. *Evolution*, 54(4), 1173-1181. <https://doi.org/10.1111/j.0014-3820.2000.tb00552.x>
- Garrido, E., Andraca-Gómez, G., & Fornoni, J. (2012). Local adaptation: simultaneously considering herbivores and their host plants. *New Phytologist*, 193(2), 445-453. <https://doi.org/10.1111/j.1469-8137.2011.03923.x>

Greischar, M. A., & Koskella, B. (2007). A synthesis of experimental work on parasite local adaptation. *Ecology letters*, 10(5), 418-434. <https://doi.org/10.1111/j.1461-0248.2007.01028.x>

Hanks, L. M., & Denno, R. F. (1994). Local adaptation in the armored scale insect *Pseudaulacaspis pentagona* (Homoptera: Diaspididae). *Ecology*, 75(8), 2301-2310. <https://doi.org/10.2307/1940885>

Hereford, J., & Winn, A. A. (2008). Limits to local adaptation in six populations of the annual plant *Diodia teres*. *New Phytologist*, 178(4), 888-896. <https://doi.org/10.1111/j.1469-8137.2008.02405.x>

Hijmans, R. J., and van Etten, Jacob. 2012. raster: Geographic analysis and modeling with raster data. R package version 2.0-12. <http://CRAN.R-project.org/package=raster>

Ismail, M., Tougeron, K., Vriamont, A., Hance, T., & Albittar, L. (2022). Thermal tolerance of the rosy apple aphid *Dysaphis plantaginea* and its parasitoids: Effect of low temperatures on some fitness activities of *Aphidius matricariae*. *Journal of Thermal Biology*, 110, 103377. <https://doi.org/10.1016/j.jtherbio.2022.103377>

Kaltz, O., & Shykoff, J. A. (1998). Local adaptation in host-parasite systems. *Heredity*, 81(4), 361. <https://doi.org/10.1046/j.1365-2540.1998.00435.x>

Kaltz, O., Gandon, S., Michalakis, Y., & Shykoff, J. A. (1999). Local maladaptation in the anther-smut fungus *Microbotryum violaceum* to its host plant *Silene latifolia*: evidence from a cross-inoculation experiment. *Evolution*, 53(2), 395-407. <https://doi.org/10.1111/j.1558-5646.1999.tb03775.x>

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology letters*, 7(12), 1225-1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>

Koupilová, K., Koubek, T., Cornille, A., & Janovský, Z. (2021). Local maladaptation of the anther-smut fungus parasitizing *Dianthus carthusianorum*. *European Journal of Plant Pathology*, 160(2), 365-374. <https://doi.org/10.1007/s10658-021-02249-0>

Miñarro, M., & Dapena, E. (2007). Resistance of apple cultivars to *Dysaphis plantaginea* (Homoptera: Aphididae): Role of tree phenology in infestation avoidance. *Environmental Entomology*, 36(5), 1206-1211. [https://doi.org/10.1603/0046-225X\(2007\)36\[1206:ROACTD\]2.0.CO;2](https://doi.org/10.1603/0046-225X(2007)36[1206:ROACTD]2.0.CO;2)

Minarro, M., & Dapena, E. (2008). Tolerance of some scab-resistant apple cultivars to the rosy apple aphid, *Dysaphis plantaginea*. *Crop Protection*, 27(3-5), 391-395. <https://doi.org/10.1016/j.cropro.2007.07.003>

Mopper, S., & Strauss, S. Y. (Eds.). (2013). *Genetic structure and local adaptation in natural insect populations: effects of ecology, life history, and behavior*. Springer Science & Business Media.

Nuismer, S. L., & Gandon, S. (2008). Moving beyond common-garden and transplant designs: insight into the causes of local adaptation in species interactions. *The American Naturalist*, 171(5), 658-668. <https://doi.org/10.1086/587077>

Nooten, S. S., & Hughes, L. (2017). The power of the transplant: direct assessment of climate change impacts. *Climatic Change*, 144, 237-255. <https://doi.org/10.1007/s10584-017-2037-6>

Olazcuaga, L., Foucaud, J., Deschamps, C., Loiseau, A., Claret, J. L., Vedovato, R., ... & Estoup, A. (2022). Rapid and transient evolution of local adaptation to seasonal host fruits in an invasive pest fly. *Evolution Letters*, 6(6), 490-505. <https://doi.org/10.1002/evl3.304>

Olvera-Vazquez, S. G., Remoué, C., Venon, A., Rousselet, A., Grandcolas, O., Azrine, M., ... & Cornille, A. (2021). Large-scale geography survey provides insights into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa. *Peer Community Journal*, 1. <https://doi.org/10.24072/pcjournal.26>

Pagliarani, G., Dapena, E., Miñarro, M., Denancé, C., et al., (2016). Fine mapping of the rosy apple aphid resistance locus Dp-fl on linkage group 8 of the apple cultivar 'Florina'. *Tree genetics & genomes*, 12(3), 56. <https://doi.org/10.1007/s11295-016-1015-x>

Selzer, G. J., Rueden, C. T., Hiner, M. C., Evans III, E. L., Harrington, K. I., & Eliceiri, K. W. (2023). napari-imagej: ImageJ ecosystem access from napari. *Nature Methods*, 1-2. <https://doi.org/10.1038/s41592-023-01990-0>

Sicard, D., Pennings, P. S., Grandclément, C., Acosta, J., Kaltz, O., & Shykoff, J. A. (2007). Specialization and local adaptation of a fungal parasite on two host plant species as revealed by two fitness traits. *Evolution*, 61(1), 27-41. <https://doi.org/10.1111/j.1558-5646.2007.00003.x>

Spitzer, B. (2006). Local maladaptation in the soft scale insect *Saissetia coffeae* (Hemiptera: Coccidae). *Evolution*, 60(9), 1859-1867. <https://doi.org/10.1111/j.0014-3820.2006.tb00529.x>

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12), 673-680. <https://doi.org/10.1016/j.tree.2014.10.004>

Warneys, R., Gaucher, M., Robert, P., Aligon, S., Anton, S., Aubourg, S., ... & Heintz, C. (2018). Acibenzolar-S-Methyl Reprograms Apple Transcriptome Toward Resistance to Rosy Apple Aphid. *Frontiers in plant science*, 9, 17. <https://doi.org/10.3389/fpls.2018.01795>

World Meteorological Organization WMO. International Organizations, 2022. Accessed September 25th 2023 [https://public.wmo.int/en/our-mandate/climate/wmo-statement-state-of-global-climate/Europe - :~:text=It provides information on rising,0.5 °C per decade.](https://public.wmo.int/en/our-mandate/climate/wmo-statement-state-of-global-climate/Europe-%20-%20text=It%20provides%20information%20on%20rising%200.5%20%C2%B0C%20per%20decade)

7.7 SUPPLEMENTARY TEXT

SUPPLEMENTARY TEXT S1

APPLE GENOTYPE SELECTION

Apples used in this study were cultivated (*Malus domestica* Borkh) and wild (*Malus sylvestris* L. Mill.) apples, with different genotypes for each. The cultivated apple genotypes were selected to represent local genotypes genetically far from each other and having variability in the response against rosy apple aphid attacks. The wild apple genotypes, we chose them because of the already-characterized population genetic differentiation that has been observed in the European wild apple (Cornille et al., 2015). We, however, acknowledge that the current experiment gave a first insight into the natural response of the wild apple genotypes to the attacks of the rosy apple aphid. The sites for settling the common garden orchards represent a European latitudinal gradient to test the effect of local environments on the rosy apple aphid adaptation.

APHID CLIP CAGE

We developed special clip cages to protect the aphids during the cross-infestation experiment. We cut 26,000 clips with foam size 5cmx5cm using a laser cutter machine (Figure 1). We collaborated with the FabLab University Paris Saclay (<https://fablabdigiscope.universite-paris-saclay.fr/>). We prepared each clip, adding a mesh on one side of the clip. We assembled each clip, placing the leaves inside of the clip cage and keeping them closed with an elastic band (Figure 1).



Figure 1. Aphid clip cage was used for the aphid cross-infestation experiment. The clip cages comprised two foam squares protected with a mesh on a single side. After placing the aphid in the leaf, we protected it with this aphid clip cage and closed it with two elastic bands.

SUPPLEMENTARY TEXT S2

The Napari plug-in pipeline was developed by the physics laboratory of Angers University led by Dr. David Rosseau and the research engineer Herarii Metuarea (Figure 2). We photographed each aphid colony and saved it in TIF format. The first step of the pipeline was the format conversion from TIF to H5 using ImageJ (Schneider et al., 2012). We used Ilastik (Sommer et al., 2011) and the included watershed algorithm for the segmentation stage. This step allowed us to separate each individual of each colony. We obtained a new set of photographs to start the classification stage. We applied a second step of conversion format from TIF to H5 using Image J. We used Napari and opened the Napari-aphid plug-in to start the classification. The aphid classification was created based on the morphology of the aphids we defined with the support of researchers, experts in the rosy apple aphid (Figure 3), Dr. Marcos Miñarro (SERIDA, Asturias, Spain) and Dr. Ahmar Ahlmedi (PC-Fruit Center, Sint-Truident, Belgium). At this point, we manually check the photographs and correct them in case of misclassification. We launch the napari-aphid plug-in and obtain the number of individuals, the different aphid stages, and the area of each individual in pixels.

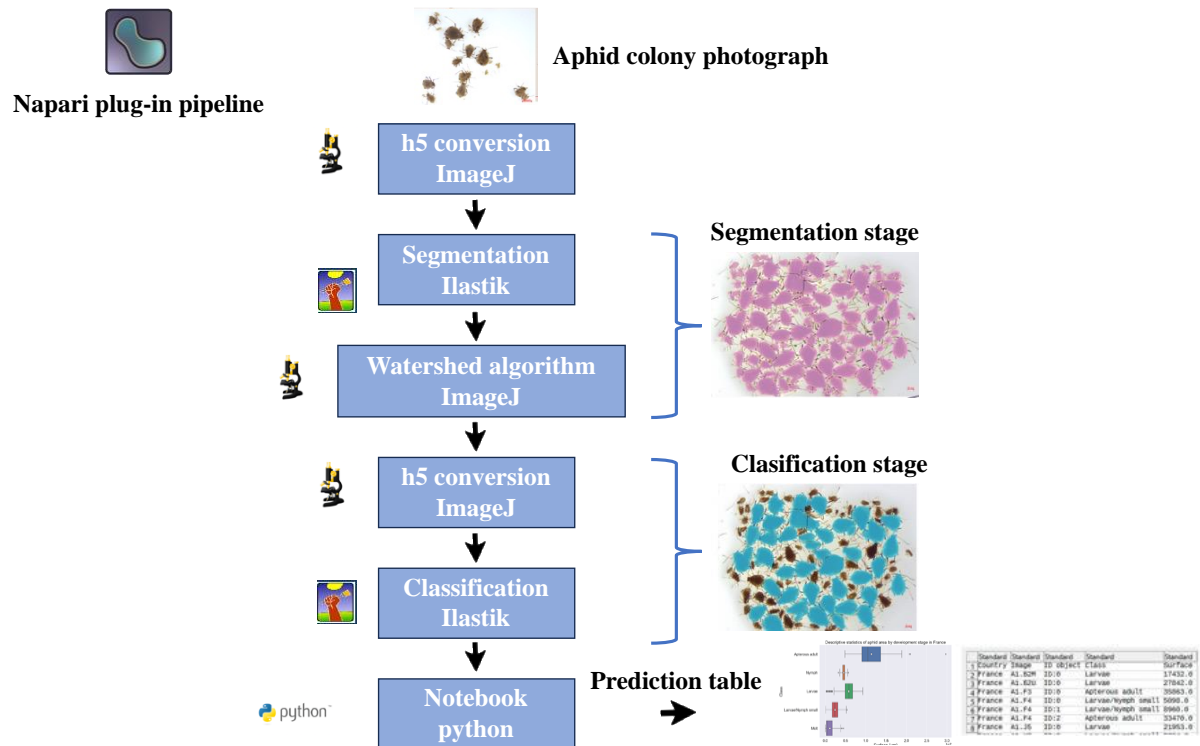
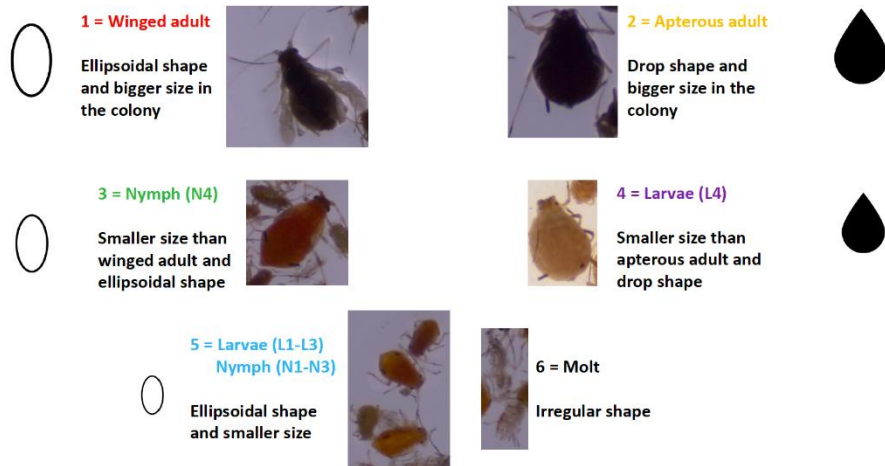


Figure 2. General napari aphid plug-in pipeline. After taking a picture of each aphid colony in TIF format, we proceeded to photograph conversion to h5 using ImageJ (Schneider et al., 2012). We started the segmentation stage using Ilastik and the watershed algorithm (Sommer et al., 2011). This step allows the separation of each individual in the colony. The next step is classifying each aphid development stage (larvae, nymph, apterous adult, winged form). We then run the napari aphid plug-in to get information about the number of individuals, the different life stages, and the surface of the individuals.

A)

Proposition for aphid classification



B)

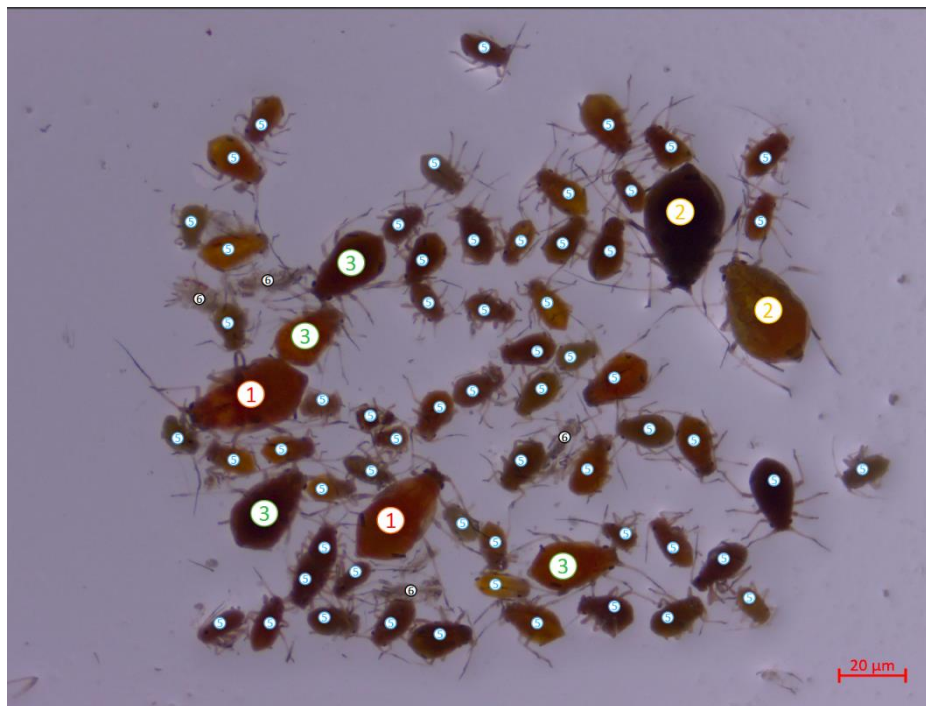


Figure 3. Aphid classification for the processing using napari aphid plug-in. This classification was created with the support of expert aphid researchers of Dr. Marcos Miñarro (SERIDA, Asturias, Spain) and Dr. Ahmar Ahlmedi (PC-Fruit Center, Sint-Truident, Belgium). Six categories were defined: 1 = winged adult, 2 = Apterous adult, 3 = Nymph larvae forms precursors of winged forms, 4 = Larvae precursors of apterous form, 5 = Larvae, 6 = molt. The aphid classification is presented on upper side (A). An example of the classification with an aphid colony is shown below (B).

REFERENCES

- Cornille, A., Feurtey, A., G elin, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., & Giraud, T. (2015). Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples. *Evolutionary applications*, 8(4), 373-384. <https://doi.org/10.1111/eva.12250>
- Napari contributors (2019). *napari: a multi-dimensional image viewer for python*. [doi:10.5281/zenodo.3555620](https://doi.org/10.5281/zenodo.3555620)
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9(7), 671-675. <https://doi.org/10.1038/nmeth.2089>
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7), 676-682. <https://doi.org/10.1038/nmeth.2019>
- Sommer, C., Straehle, C., Koethe, U., & Hamprecht, F. A. (2011, March). Ilastik: Interactive learning and segmentation toolkit. In *2011 IEEE international symposium on biomedical imaging: From nano to macro* (pp. 230-233). IEEE. [10.1109/ISBI.2011.5872394](https://doi.org/10.1109/ISBI.2011.5872394)
- Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

7.8 SUPPLEMENTARY MATERIAL

Table S1. Country of origin, species, name or identification ID, number of trees, assigned to each common garden. B = Belgium, F = France, and S = Spain. Each genotype has an identification including 1) the genotype name and 2) the accession ID.											
Origin of the genotypes	ID	Common garden orchards			TOTAL	Origin of the genotypes	ID	Common garden orchards			TOTAL
		B	F	S				B	F	S	
Belgium (<i>Malus domestica</i>)	Braeburn_P03a01	12	11	12	173	European wild apple Belgium (<i>Malus sylvestris</i>)	syl_be 148	10	10	10	197
	Elstar_P03a02	12	11	12			syl_be 4	11	11	12	
	Fuji_P03a12	11	11	12			syl_be 54	11	10	11	
	Granny Smith_P03a04	12	11	12			syl_be 60	11	10	11	
	Wellant_V05a1	11	11	12			syl_be 76	12	11	12	
	Total Belgian trees	58	55	60			Total Belgian wild apple trees	66	63	68	
France (<i>Malus domestica</i>)	Api_Noir_	12	11	12	173	European wild apple Spain (<i>Malus sylvestris</i>)	syl_es B	11	11	11	97
	Clochard_A5	12	11	12			syl_es D	10	9	10	
	Reale_d'Entraygues	11	11	11			syl_es F	12	11	12	
	Reinette_Franche	12	11	12			Total Spanish wild apple trees	33	31	33	
	Reine Des Reinettes Tasse	12	11	12			Total European wild apple trees	99	96	101	
Total French trees	59	55	59								
Spain (<i>Malus domestica</i>)	Limón_Montés_M0236	12	11	12	173	Tolerant control	<i>Malus floribunda</i> _X6518	11	11	11	100
	Perico_M0056	11	11	12			Florina_X2775	11	10	11	
	Raxao_M0174	12	11	12			Priscilla X2851	12	11	12	
	Regona_M0239	11	11	12		Total per site (for infestations: modality 1)	244	220	246	710	
	Xuanina_M0084	12	11	12		Total per site (control without infestations: modality 2)	76	83	80	239	
	Total Spanish trees	58	55	60							
Susceptible control	Golden Delicious cv.	12	12	12	36		Total	320	305	326	951
						Aphid rearing and synchronization (February 2021)	Golden Delicious cv.	63	80	63	206
						TOTAL over sites	(infestation + rearing)				1193 trees

Table S2. Description of the indexes, terms, and effects included in our proposed statistical model General Linear Mixed Model (GLMM) to test for local adaptation of the rosy apple aphid (*Dysaphis plantaginea*).

Math index	Index	Term	Effect
α	h	Aphid_origin _h	Aphid country of origin (Spain, France, Belgium), fixed effect
β	i	Apple_origin _i	Apple country of origin (Spain, France, Belgium), fixed effect
γ	j	Site _j	Common garden site (Spain, France, Belgium), with a covariance-variance matrix of difference of temperature (or humidity) between each apple tree before (or after) the infestation, fixed effect
B	k	Block _k	Block (each block consists of 28 apple genotypes infested with 9 aphid genotypes), random effect
	l	Gh _l	Apple host genotype, random effect
μ	m	Leaf_level _m	Leaf level (Position of the infested apple leaf on the main stem. Three levels: upper, middle, or lower), random effect
P	n	Gp _n	Aphid parasite genotype, random effect
ω	o	Tree_clone _o	Apple clone of a given genotype, random effect
δ	t	Time of infestation _t	Day of infestation, random effect
ζ	t2	Time of infestation _{t2}	Hour of infestation, random effect
	z		Effect of each observation
ε			Residual error

Note that the Belgium site and genotypes were removed from the statistical analyses because of a storm that occurred in 2021.

Table S3. Summary of Model 1 within the Spanish common garden (*i.e.*, removing common garden by aphid origin interaction effect, GLMM, Negative Binomial). *P-values* are explained below in the Table (*: $P < 0.05$ **: $P < 0.001$). The variable response was the colony growth rate. Origin_aphid: origin of the aphid genotype, French or Spanish; Count_origin_tree: country of origin of the apple genotypes, France or Spain, resistant or tolerant.

ANOVA Response: Colony growth rate Spanish site			
	X^2	D.F.	Pr> X^2
Origin_aphids	0.031	1	0.860
Count_origin_tree	0.831	4	0.934

Table S4. Summary of Model 1 within the French common garden (*i.e.*, removing common garden by aphid origin interaction effect, GLMM, Negative Binomial). *P-values* are explained below in the Table (*: $P < 0.05$ **: $P < 0.001$). The variable response was the colony growth rate. Origin_aphid: origin of the aphid genotype, French or Spanish; Count_origin_tree: country of origin of the apple genotypes, France or Spain, resistant or tolerant.

ANOVA Response: Colony growth rate French site				
	X ²	D.F.	Pr>X ²	
Origin_aphids	5.746	1	0.016*	
Count_origin_tree	3.807	4	0.433	

Table 5. Summary of Model 1 within the Spanish aphid population (*i.e.*, removing common garden by aphid origin interaction effect, GLMM, Negative Binomial). *P-values* are explained below in the Table (*: $P < 0.05$ **: $P < 0.001$). The variable response was the colony growth rate. Common_garden: common garden, French or Spanish; Country_origin_tree: country of origin of the apple genotypes, France or Spain, resistant or tolerant.

ANOVA Response: Colony growth rate Spanish aphids			
	X^2	D.F.	$Pr > X^2$
Common_garden	1.810	1	0.178
Count_origin_tree	1.486	4	0.829
Common_garden:count_origin_tree	1.592	4	0.810

Table S6. Summary of Model 1 within the French aphid population (*i.e.*, removing common garden by aphid origin interaction effect, GLMM, Negative Binomial). *P-values* are explained below in the Table(*: $P < 0.05$ **: $P < 0.001$). The variable response was the colony growth rate. Common_garden: common garden French or Spanish; Country_origin_tree: country of origin of the apple genotypes, France or Spain, resistant or tolerant.

Response: Colony growth rate French aphids			
	X ²	D.F.	Pr>X ²
Common_garden	12.857	1	0.001***
Count_origin_tree	0.873	4	0.928
Common_garden:count_origin_tree	4.767	4	0.312

7.9 ANNEXES: EXPERIMENTAL TEST FOR LOCAL ADAPTATION OF THE ROSY APPLE APHID (*DYSAPHIS PLANTAGINEA*) DURING ITS RECENT PAID COLONIZATION ON ITS CULTIVATED APPLE HOST (*MALUS DOMESTICA*) IN EUROPE



PREREGISTRATION ARTICLE



Open Access



Open Peer-Review

Cite as: Olvera-Vazquez S.G., Alhmedi A., Miñarro M., Shykoff J. A., Marchadier E., Rousselet A., Remoué C., Gardet R., Degrave A., Robert P., Chen X., Porcher J., Giraud T., Vander-Mijnsbrugge K., Raffoux X., Falque M., Alins, G., Didelot F., Beliën T., Dapena E., Lemarquand A. and Cornille A. (2021) Experimental test for local adaptation of the rosy apple aphid (*Dysaphis plantaginea*) to its host (*Malus domestica*) and to its climate in Europe. In principle recommendation by Peer Community In Ecology. https://forgemia.inra.fr/mandine.cornille/local_adaptation_dp.ver.4.

Posted: 23rd April 2021

Recommender: Eric Petit

Reviewers: Alex Stemmelen, Sharon Zytynska and one anonymous reviewer

Correspondence:
amandine.cornille@inrae.fr

Experimental test for local adaptation of the rosy apple aphid (*Dysaphis plantaginea*) during its recent rapid colonization on its cultivated apple host (*Malus domestica*) in Europe

Olvera-Vazquez S.G.¹, Alhmedi A.², Miñarro M.³, Shykoff J. A.⁴, Marchadier E.¹, Rousselet A.¹, Remoué C.¹, Gardet R.⁵, Degrave A.⁵, Robert P.⁵, Chen X.¹, Porcher J.⁵, Vander-Mijnsbrugge K.⁶, Raffoux X.¹, Falque M.¹, Anaya-Sainz J.M.¹, Deldycke K.¹, Gay R.¹, Alins, G.⁷, Giraud T.³, Didelot F.⁸, Beliën T.², Dapena E.³, Lemarquand A.⁸, Cornille A.¹

¹ Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190 Gif-sur-Yvette, France

² Department of Zoology, pcfuit vzw, Sint-Truiden, Belgium

³ Servicio Regional de Investigación y Desarrollo Agroalimentario (SERIDA), Carretera AS-267, PK. 19, E-33300, Villaviciosa, Asturias, Spain

⁴ Laboratoire d'Ecologie, Systématique et Evolution, Université Paris-Saclay, CNRS, AgroParisTech, 91400 Orsay cedex, France.

⁵ AgroCampus-Ouest, UMR1345 Institut de Recherche en Horticulture et Semences (IRHS), 49045 Angers, France

⁶ Department of Forest Ecology and Management, Research Institute for Nature and Forest, 9500 Geraardsbergen, Belgium

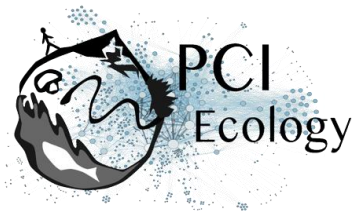
⁷ IRTA Fruitcentre, PCiTAL, Park of Gardeny, Fruitcentre Building, 25003 Lleida, Spain

⁸ INRAE, Unité Expérimentale Horticole N34 0449, Centre d'Angers-Nantes, 49071 Beaucozé Cedex, France

This preregistration article has been peer-reviewed and recommended by *Peer Community in Ecology* <https://doi.org/10.24072/pci.ecology.100079>

ABSTRACT

Understanding the extent of local adaptation in natural populations and the mechanisms enabling populations to adapt to their environment is a major avenue in ecology research. Host-parasite interaction is widely seen as a major driver of local adaptation and has therefore been a study model to dissect the evolutionary processes at work during local adaptation. However, to date, the relative contributions of species interactions (*i.e.*, biotic factor) and abiotic factors to local adaptation are still unclear. Addressing these issues is more than a simple academic exercise. Understanding local adaptation processes in host-parasite interactions will also help to tackle pressing issues, such as the ways in which environmental changes alter the emergence of pathogens leading to host extinction, how to promote sustainability of agroecosystems in the face of emerging crop diseases, or in guiding public health practices



as more human pathogens and their vectors expand their ranges. Here, we propose to investigate whether local adaptation occurred during the recent rapid colonization of cultivated apple (*Malus domestica*) by *Dysaphis plantaginea*, the major aphid pest of cultivated apple orchards in Europe. We will experimentally test whether different populations, from Belgium, France, and Spain, of the aphid *D. plantaginea* show fitness differences in three common garden orchards located in Belgium, France, and Spain, comprised each of a panel of wild and cultivated apple genotypes from Belgium, France, and Spain, as well as previously reported tolerant and susceptible apple genotypes. This experiment will start in the Spring of 2021 and will generate original results adding to our understanding of how the biotic (the host) and abiotic conditions can shape local adaptation in a parasite.

Keywords: local adaptation, aphid, fruit trees, apple, common garden, G*G*E interaction, host-parasite interaction, domestication

Research question

Hypotheses

The general question that we would like to address is whether there is a pattern of local adaptation of the rosy apple aphid (*Dysaphis plantaginea* Passerini) to 1) its local environment and/or 2) its cultivated apple host (*Malus domestica* Borkh)? To that aim, we will answer several questions, outlined below.

Question 1 and hypotheses: Is there evidence of rosy apple aphid adaptation to the local environment? Note here that the local environment will be tested with the “site” effect (Equation 1), which includes abiotic (*i.e.*, soil or climate) and biotic (*i.e.*, other aphid species and parasites of the cultivated apple host) factors. However, the biotic effect of the local cultivated apple host will be tested separately in Question 2. Note also that we will record the temperature of each leaf before and after the infestation. This temperature record per leaf will be used for statistical analyses to specifically test whether temperature plays a role in aphid infestation success (see statistical analyses part).

Do the rosy apple aphid genotypes from three different origins (Belgium, France, and Spain) show higher fitness in their local environment (*i.e.*, Belgium, France, and Spain, respectively) and lower fitness in their foreign environment (Figure 1)?

Hypothesis 0: There are no differences among the aphid populations across the three common garden orchards (Belgium, France, and Spain).

Hypothesis 1: There are differences among the aphid populations across the three common gardens (Belgium, France, and Spain). A significantly higher aphid fitness in the local common garden, while lower elsewhere, will support the hypothesis of local adaptation of the rosy apple aphid to its local environment. A significantly lower aphid fitness in the local common garden, while higher elsewhere, will support the hypothesis of maladaptation (Capblancq et al., 2020). Local adaptation of parasites is not a universal phenomenon; maladaptation has been observed in some systems such as the obligate parasite *M. violaceum* on its host *Silene latifolia* Poir. (Kaltz et al., 1999; Koupilová et al., 2021), with higher resistance of sympatric hosts. For aphids, only a handful of studies have been performed to test for local adaptation of aphids, and only to their hosts (Smadja et al., 2012; Simon et al., 2015; Biello et al., 2021).

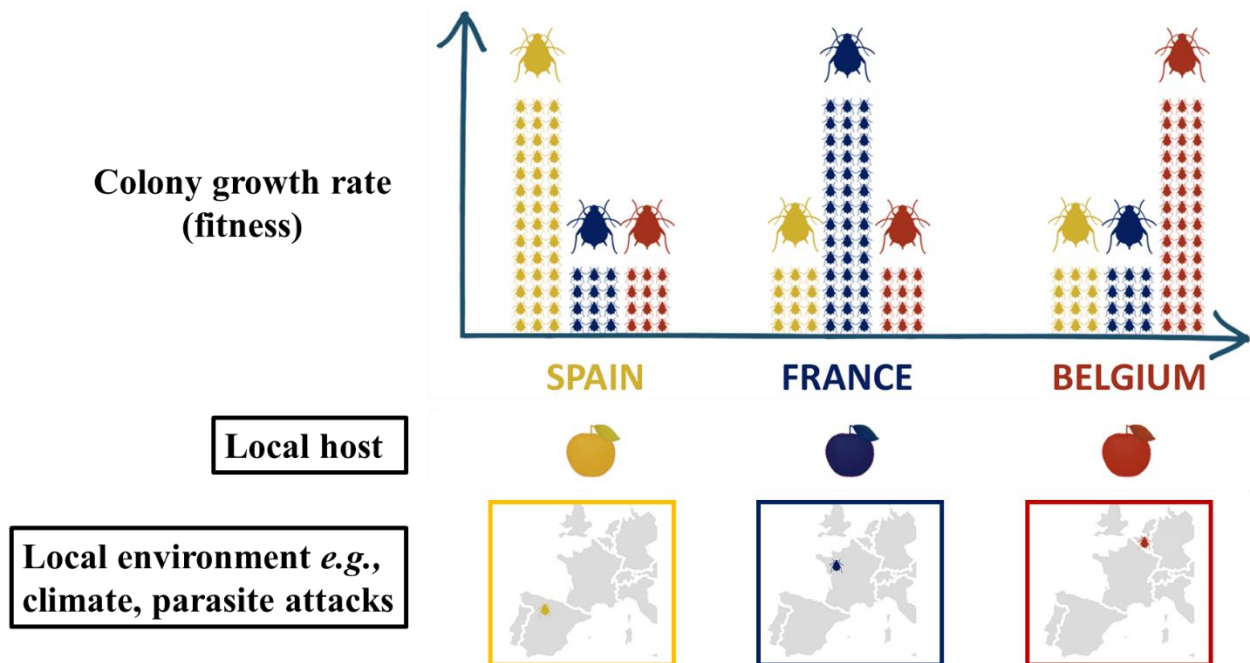


Figure 1. Expected patterns in the case of the rosy apple aphid (*Dysaphis plantaginea*) are locally adapted to its local environment and host. The rosy apple aphid populations that present the highest fitness in their local abiotic environment and host will reflect local adaptation.

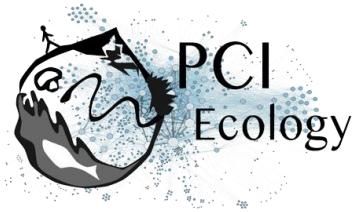
Question 2 and hypotheses: Is there evidence of rosy apple aphid adaptation to the local cultivated apple host genotypes?

Do the rosy apple aphid genotypes from three different origins (*i.e.*, Belgium, France, and Spain) show higher fitness on their respective local apple host genotypes (*i.e.*, local Belgian, French, and Spanish apple genotypes, respectively) and lower fitness on their foreign apple genotypes (Figure 1)?

Hypothesis 0: There are no differences among the aphid populations infested on the different local apple genotypes from different origins (Belgium, France, and Spain).

Hypothesis 1: There are differences among the aphid populations infested on the different apple genotypes from different origins (Belgium, France, and Spain). A significantly higher aphid fitness on local apple genotypes, while lower on non-local apple genotypes, will support the hypothesis of local adaptation of the rosy apple aphid to its host. A significantly lower aphid fitness on local apple genotypes, while higher on all other apple genotypes, will support the hypothesis of maladaptation of the rosy apple aphid to its host.

Question 3 and hypotheses: Is there evidence of rosy apple aphid adaptation to the local cultivated apple host and the local environment?



Is the fitness of the rosy apple aphid genotypes from three different origins (Belgium, France, and Spain) higher on their respective local cultivated apple host (*i.e.*, local Belgian, French, and Spanish apple genotypes) and in their respective local environment (*i.e.*, local Belgian, French, and Spanish), compared with the fitness of the different rosy apple aphid genotypes on foreign apple host genotypes and the foreign environment (Figure 1)?

Hypothesis 0: There are no differences among the aphid populations infested on the different local apple genotypes from different origins (Belgium, France, and Spain) and across the three common gardens (Belgium, France, and Spain).

Hypothesis 1: There are differences among the aphid populations infested on the different local apple genotypes from different origins (Belgium, France, and Spain) and across the three common gardens (Belgium, France, and Spain). A significantly higher aphid fitness on the local apple genotypes and at the local common garden, while lower elsewhere, will support the hypothesis of local adaptation of the rosy apple aphid to its environment and host. On the other hand, a significantly lower aphid fitness on the local apple genotypes, and at the local common garden, while higher elsewhere, will support the hypothesis of maladaptation.

Question 4 and hypotheses: Is the rosy apple aphid adapted to the cultivated apple or to the locally occurring wild apple in Europe?

Is the fitness of the rosy apple aphid higher on the cultivated apple host than on the European wild apple *Malus sylvestris* (L.) Mill?

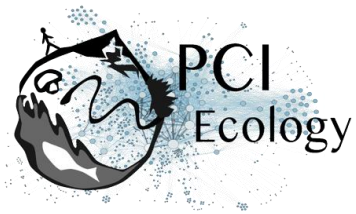
Hypothesis 0: There are no differences among the aphid populations infested either on wild or cultivated apple genotypes.

Hypothesis 1: There are fitness differences among the aphid populations infested on wild apple genotypes and cultivated apple genotypes. A significantly higher aphid fitness on the wild apple genotypes will support the hypothesis that the rosy apple aphid is better adapted to the local wild apple. So far, there is no information on how the domestication of the apple tree could have altered resistance to aphid infestation, but we can suggest a hypothesis. The European wild apple is the local genotype in Europe and has been present there for at least the past 120,000 years. The cultivated apple has less time in Europe; it was brought by the Romans and Greeks in Europe about 1,500 years ago (Cornille et al. 2014, 2019). Current population genetics evidence (Olvera-Vazquez et al. 2020) along with the rosy apple aphid geographic distribution (mainly in Europe and the Middle East) suggest a long-time association with the European wild apple than the cultivated apple. As a result, the rosy apple aphid had more time to adapt to the European wild apple.

Question 5 and hypotheses:

Is the fitness of the rosy apple aphid lower on apple genotypes known *a priori* to be tolerant (Pagliarani et al., 2016, Marchetti et al. 2009) to the rosy apple aphid?

Hypothesis 0: There are no differences among the aphid populations infested on apple genotypes known to be tolerant to the rosy apple aphid infestation and on other susceptible apple genotypes.



Hypothesis 1: Yes, there are fitness differences between aphids infested on the tolerant apple genotypes (*Malus floribunda* Siebold ex Van Houtte, *M. domestica* Florina, and *M. domestica* Priscila) and the susceptible apple genotype (*M. domestica* Golden Delicious). Previous studies suggested that the apple genotypes tolerant to the rosy apple aphid infestations induced lower fitness of the rosy apple aphid (Miñarro and Dapena, 2007; Pagliarani et al., 2016; Dall'Agata et al., 2018).

Sampling plan

Overall design

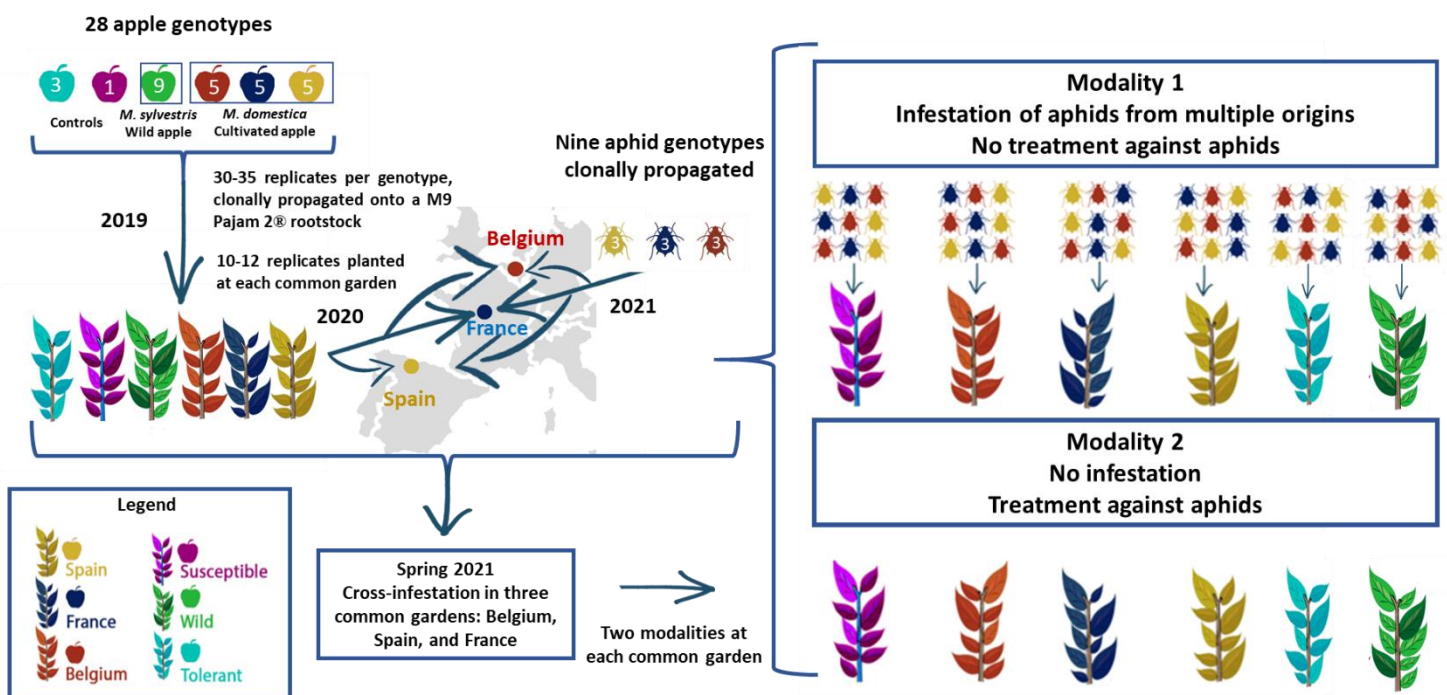
The experiment will be located at three common garden orchards at 1) Sint-Truiden in Belgium (50°48'0" N, 5° 11'0" E), presenting a mean annual temperature of 9.6°C and annual precipitation of 823 mm, 2) Beaucozé in France (47°28'57" N, 0°36'52" W), presenting a mean annual temperature of 11.4°C and annual precipitation of 675 mm annual precipitation, and 3) Villaviciosa in Asturias in Spain (43°28'45" N, 5° 26'32" W), presenting a mean annual temperature of 11.8°C and annual precipitation of 869 mm. The bioclimatic information was extracted from the WorldClim – Global Climate database <https://www.worldclim.org/> (Fick et al., 2017) with the raster R package (Hijmans and van Etter, 2012). In the spring of 2021, we will perform an infestation experiment using nine aphid genotypes, each representing the clonal offspring of a single female (called hereafter matriline) that had been collected in Belgium, France, and Spain, with three matriline from each country. Below we describe the material that will be used.

Apple trees

Each common garden is made of 28 apple genotypes (Figure 2, Table 1). A total of 15 cultivated apple genotypes (*M. domestica*) comes from three countries, with five local genotypes from each country. The selection of the local cultivated apple genotypes was based on several criteria. First, whenever possible the genotypes were chosen to be apple genotypes locally cultivated in the surrounding area of each common garden. In the cases of Spain and France, the local genotypes encompass traditional genotypes, while in Belgium, the cultivation of apple encloses recent commercial genotypes. Second, we chose cultivated genotypes inferred not to be the most genetically closely related based on microsatellite genetic characterization (Cornille et al., 2012). Third, unpublished qualitative assessments of *D. plantaginea* attacks onto several cultivated apple varieties allowed choosing five apple varieties per locality that showed variability in their response to *D. plantaginea* infestation (from susceptible to tolerant). We also added nine wild apple genotypes (*M. sylvestris*), six from Belgium, and three from Spain. We obtained scions from mother trees maintained in a conservation orchard in Belgium, and from sampling in a forest in Northern Spain. The choice of the genotypes was based on previous studies that showed that Spanish and Belgian wild apples belonged to genetically differentiated populations in Europe (Cornille et al. 2013, 2015). Note that we failed to obtain scions for French wild apple genotypes in the year of the grafting. We also included four apple genotypes with different susceptibility levels to aphid infestations: three tolerant apple genotypes (two *M. domestica* apple

genotypes, ‘Priscilla’ and ‘Florina’ genotypes, and one genotype of the ornamental species *Malus floribunda*), and one susceptible genotype, the *M. domestica* Golden Delicious genotype. We selected these apple genotypes to have a range of tolerance to *D. plantaginea* infestation (Miñarro and Dapena, 2007; Pagliarini et al., 2016). Note that the 28 apple genotypes used in this experiment have been genetically characterized using 13 microsatellite markers (data not shown), and we sequenced their genomes (Illumina sequencing), which will be analyzed in 2021.

According to the availability of the scions at the beginning of the project in 2018, we grafted 10 to 12 clonemates for each of the 28 apple genotypes (Figure 2, Table 1). Besides, for the aphid rearing and synchronization steps that will be performed at each common garden orchard (see method below), we also grafted 206 clonemates of the Golden Delicious genotype (Table 1), to get at least 60 trees per locality available for the rearing. In total, 1,157 apple trees (Table 1, 951 for the infestation experiment and 206 for the rearing step) were grafted in early 2019 on an M9 Pajam 2® apple rootstock and maintained for one year (February 2019-2020) at an outdoor nursery at La Retuzière, Beaucouzé, Angers, France (47°28’57” N, 0°36’52” W). In early February 2020, the trees were transferred and planted in the three common garden orchards (Figure 2). Each tree was sprayed with Teppeki® (flonicamida 50%) insecticide, a Bordeaux mixture (20% copper) fungicide, DELFIN® (*Bacillus thuringiensis* sp. *kurstaki*) anti-lepidopterous, Essen’ciel (orange essential oil) insecticide and fungicide, Karate Zeon® (Lambda cihalotrin 1.5%) and Movento® (Spirotetramat 15% p/v OD) insecticides, and Sokalcarbio WP® (calcined kaolin), a mineral physical barrier between pest and plants. These treatments will be continued until the beginning of the experiment (March 2021). We will also apply an aphicide and fungicide treatment two weeks before the beginning of the aphid infestation experiment (Figure 2).



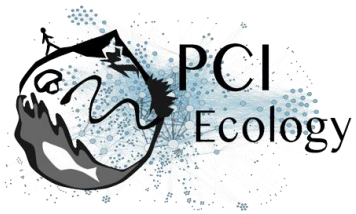


Figure 2. General scheme of the aphid cross-infestation experiment that will be performed in the Spring of 2021 at the three common garden orchards in Belgium, Spain, and France. At each common garden orchard, 28 clonally propagated apple genotypes are grown with 10 to 12 replicates per genotype, depending on the survival of the grafted trees at each common garden. The apple genotypes included 1) *Malus domestica* genotypes from Belgium (five genotypes, red color), France (five genotypes, dark blue color), and Spain (five genotypes, yellow color). Additionally, 2) nine wild apple genotypes (*Malus sylvestris*), including six from Belgium and three from Spain (light green color), 3) *M. domestica* genotypes (Priscilla and Florina cultivars), and *Malus floribunda* Siebold ex Van Houtte, used as “tolerant to aphid infestation” controls (light blue color), and 4) the Golden delicious *M. domestica* genotype that will be used for aphid rearing as well as the “susceptible to aphid infestation” control (purple). Meanwhile, nine rosy apple aphid genotypes (*Dysaphis plantaginea*) were clonally propagated: three from Belgium (red color), three from France (dark blue color), and three from Spain (yellow color). A total of 10-12 replicates of each of the 28 apple genotypes were transferred in February 2020 to each of the three common gardens. The aphid genotypes will be transferred for rearing locally in February 2021 at each site.

Rosy apple aphid genotypes

We collected 36 rosy apple aphid colonies on several cultivated apple trees at each common garden during the spring of 2020. These aphid colonies were 12 colonies from Belgium, eight colonies from France, and 16 colonies from Spain. The colonies were sent to the GQE-Le Moulon laboratory at University Paris-Saclay in France. The colonies, consisting of one to several genotypes, are currently being reared and maintained asexually in a climate chamber at 20°C, 60-65% of relative humidity, 16 hours of light, and 8 hours of dark) on *in vitro* apple plants (Jonagold genotype) provided by the CRA-W (Micropropagation laboratory, Biological Engineering Unit, Gembloux, Belgium), in preparation for the cross-infestation experiment. The apple genotype used for aphid rearing (this case the Jonagold) was chosen to be different from any cultivars that are in the infestation experiment of the Spring of 2021, to avoid any aphid acclimatization to a specific apple genotype.

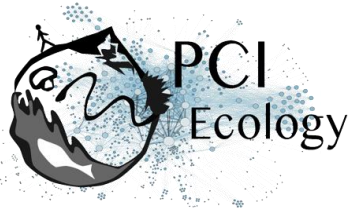
Currently, we are isolating one female from each colony onto a new *in vitro* Jonagold apple plant to ensure that we will have “single-genotype” colonies (*i.e.*, matriline) for the infestation in March 2021. Indeed, although the aphid colonies were collected to avoid mixing several clonal lineages, this can happen. Therefore, once grown up enough (about 30 individuals), we will utilize a single adult aphid to start a new colony. After the colony grows about 30-40 individuals, the colony will be genetically characterized using newly developed microsatellite markers (Olvera-Vazquez et al., 2020). This step will allow us to build a collection of at least three distinct matrilineages from each locality (*i.e.*, Belgium, France, Spain) that will be available for the infestation experiment in March 2021. Because some lines could be lost, we will maintain more than three genotypes per locality until March of 2021 in controlled conditions. In the end, from our complete set of 36 rosy apple colonies, we will maintain at least nine matrilineages from Belgium, France, and Spain. In March 2021, some progeny of each of the nine-matriline rosy apple aphid colonies will be sent to each local laboratory in Belgium, France, and Spain. Locally, each lab will rear and synchronize each of the nine colonies in a greenhouse onto Golden Delicious genotypes (63 trees in Belgium, 80 trees in France, and 63 trees for Spain; Table 1) for the infestation experiment that will be performed in March 2021.



Peer Community In Ecology

Table 1. Country of origin, species, name or identification ID, number of trees, assigned to each common garden. B = Belgium, F = France, and S = Spain. Each genotype has an identification including 1) the genotype name and 2) the accession ID.

Origin of the genotypes	ID	Common garden orchards			TOTAL	Origin of the genotypes	ID	Common garden orchards			TOTAL
		B	F	S				B	F	S	
Belgium (<i>Malus domestica</i>)	Braeburn_ P03a01	12	11	12	173	European wild apple Belgium (<i>Malus sylvestris</i>)	syl_be 148	10	10	10	197
	Elstar_ P03a02	12	11	12			syl_be 4	11	11	12	
	Fuji_ P03a12	11	11	12			syl_be 54	11	10	11	
	Granny Smith_ P03a04	12	11	12			syl_be 60	11	10	11	
	Wellant_ V05a1	11	11	12			syl_be 76	12	11	12	
	Total Belgian trees	58	55	60			Total Belgian wild apple trees	66	63	68	
France (<i>Malus domestica</i>)	Api_Noir_	12	11	12	173	European wild apple Spain (<i>Malus sylvestris</i>)	syl_es B	11	11	11	97
	Clochard_A5	12	11	12			syl_es D	10	9	10	
	Reale_d'Entraygues	11	11	11			syl_es F	12	11	12	
	Reinette_Franche	12	11	12			Total Spanish wild apple trees	33	31	33	
	Reine Des Reinettes	12	11	12			Total European wild apple trees	99	96	101	
	Tasse	12	11	12							
Total French trees	59	55	59								
Spain (<i>Malus domestica</i>)	Limón_Montés_M0236	12	11	12	100	Tolerant control	<i>Malus floribunda</i> _X6518	11	11	11	
	Perico_M0056	11	11	12			Florina_X2775	11	10	11	
	Raxao_M0174	12	11	12			Priscilla X2851	12	11	12	
	Regona_M0239	11	11	12			Total tolerant trees	34	32	34	



	Xuanina_M0084	12	11	12		Total per site (for infestations: modality 1)	244	220	246	710	
	Total Spanish trees	58	55	60	173	Total per site (control without infestations: modality 2)	76	83	80	239	
Susceptible control	Golden Delicious cv.	12	12	12	36	Total	320	305	326	951	
						Aphid rearing and synchronization (February 2021)	Golden Delicious cv.	63	80	63	206
						TOTAL over sites	(infestation + rearing)				1193 trees

Sampling size

Global design and sampling size

Each common garden orchard contains 10 to 12 clones of each of the 28 apple genotypes (Table 1). These are planted in 10 to 12 rows, each row comprised of the available genotypes placed at random (Figure 3). The experiment will be divided into two modalities (Figure 3):

- modality 1: apple genotypes infested by the rosy apple aphid genotypes from different origins; seven to nine replicates of the 28 genotypes.
- modality 2: apple genotypes free of rosy apple aphid infestations; three replicates of the 28 genotypes that will be used as non-infested controls.

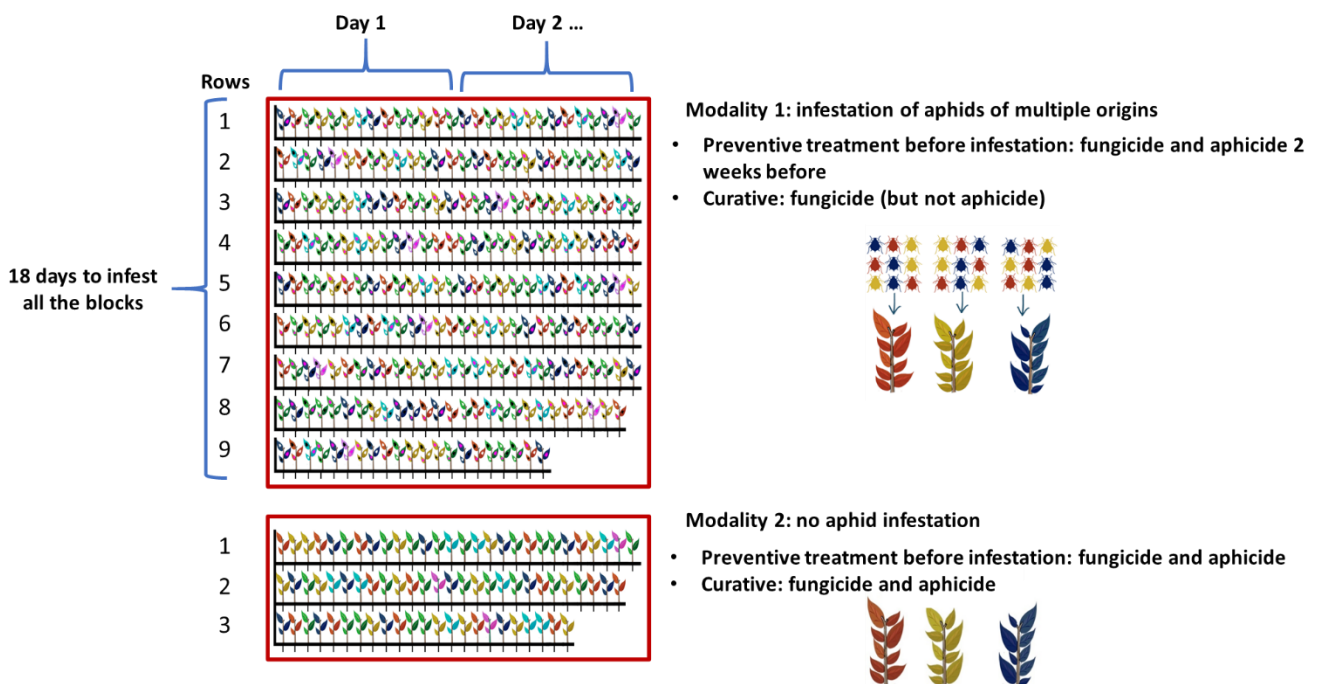


Figure 3. Details of the two modalities that will be performed during the spring of 2021 for testing the local adaptation of the rosy apple aphid (*Dysaphis plantaginea*) using a cross-infestation experiment. Here an example of the common garden in Belgium (Sint-Truiden). The experimental field of each common garden consists of rows, each including the 28 apple genotypes positioned at random in the row; the final rows lack a few genotypes due to the death of certain apple genotypes in 2019 and 2020. All trees will receive an aphicide and fungicide treatment two weeks before the infestation begins. Nine different aphid genotypes from each of the three locations (three from Belgium, three from France, and three from Spain) will then be infested on

the 28 apple trees (five genotypes from Belgium, five from France, and five from Spain, six European wild apple *M. sylvestris* genotypes, three tolerant controls, and one susceptible cultivated apple control) in mid-April 2021. Modality 1 will consist of the infestation of as many apple trees as possible per day, but we think we will need about 18 days to complete the infestation of all trees. We aimed to infest 14 apple trees as the minimal number of infested trees per day. For modality 2 (control), there will no infestation and we will apply treatments against aphids and fungi. Different colors of aphids and trees represent different genotypes. Apple trees and aphid genotypes will be spatially randomized in each block.

In the spring of 2021, we will perform a cross-infestation experiment. At that time, the planted apple genotypes will be two years old, having acclimatized to their field conditions in the common garden for one year. Each of the nine rosy apple aphid genotypes will be placed on a different leaf on the same apple tree of each of the 28 different apple genotypes in the three common garden orchards (Figures 2, 3, and 4 and Tables 1 and 2). The infestation will be performed at the apple phenological stage E2 when the development of the inflorescences occurs (Bloesch et al., 2013; Figure 4). Aphid genotypes will be placed on the leaves at random for each level of the tree (upper, middle, lower). Performing the infestation is delicate and time-consuming and will, therefore, require several days to be completed (we estimate 18 days per orchard, see Figure 3). Every day, we will record the date of initiation of each infestation and include these in the analyses as temporal blocks and the time within the days as a covariate.

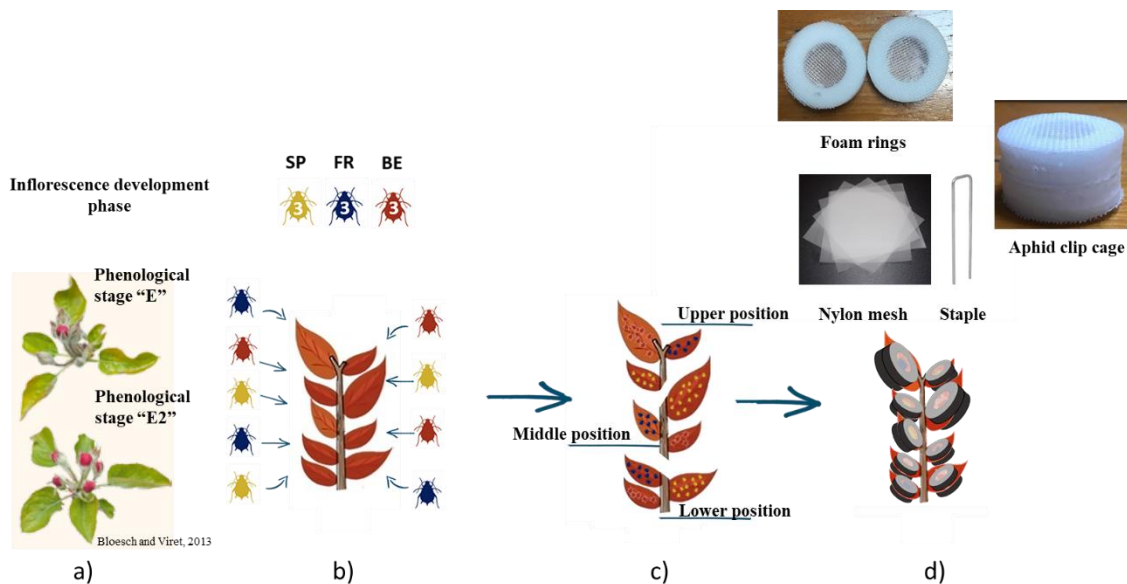
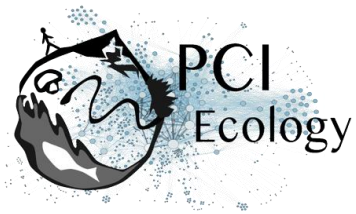


Figure 4. Representation of rosy apple aphid infestation on the different apple genotypes. a) The aphid infestation will be performed at the phenological stages “E” and “E2”. During the E stage the sepals open slightly, the petals lengthen and become visible while in the E2 stage the flowers form a hollow balloon with their petals (Bloesch et al., 2013). During both stages, there are tender light green leaves. b) Nine aphid genotypes from different origins (three from Belgium, three from France, and three from Spain) will be infested on an apple tree. c) Synchronized micro-colonies of female aphids from the nine aphid genotypes will be infested on a leaf of a tree: three aphid genotypes from France, Belgium, and Spain will be randomly infested



in the upper part of the tree, three in the medium and three in the lower part of the tree. d) Each infestation will be protected with a clip-cage. The clip cage is comprised of two circular plastazote foam rings (each ring 25 mm diameter and 1cm thickness) covered by a nylon screen and clip together with an angle-shaped staple. BE = Belgium, FR = France, SP = Spain.

In total, we plan to perform 6,408 aphid infestations on 712 apple trees across the three common gardens in Belgium, France, and Spain (Figure 3 and Table 1), with nine aphid genotypes per tree (three aphid genotypes per location, from Belgium, France, and Spain). On those trees, we will have 2,196 infestations on 244 apple trees in Belgium, 2,214 infestations on 246 trees in Spain, and 1,998 infestations on 222 trees in France (Tables 1 and 2). We expect all trees to survive, but tree sample sizes may be reduced at the start of the experiment if trees die during the fall of 2020. Overall, each aphid genotype will be confronted with 1) five cultivated apple genotypes from its native range, 2) 10 cultivated apple genotypes from two different non-native ranges, 3) nine wild apple genotypes, and 4) three apple genotypes tolerant to rosy apple aphid infestations (two *M. domestica* and one *M. floribunda*). In addition, each aphid genotype will experience the climatic conditions from its native origin and two different local environmental conditions (including abiotic and biotic factors, such as climate or soil composition, and attacks of local parasites, respectively). This will allow us to experimentally test the existence of local adaptation of the rosy apple aphid to the cultivated apple host and its local environment, as well as to compare aphid performance on wild apple (*M. sylvestris*) and on apple genotypes tolerant to rosy apple aphid infestations.

Aphid genotypes and preparation for infestation

In early March 2021, each colony will be sent from the GQE-Le Moulon laboratory to each local laboratory in Spain, France, and Belgium for aphid rearing and synchronization in local greenhouses at 20°C and 60 to 65% of relative humidity. Each colony will be reared and maintained on Golden Delicious apple trees grafted onto an M9 Pajam2® rootstock. Those Golden Delicious trees were produced at the same time as the trees used in the common gardens (*i.e.*, 2019, Table 1).

We will infest on each leaf of a tree a “mini-colony”, including two adult females and five larvae of each aphid genotype. Indeed, infesting only one female is too risky, several trials in the lab showed that infestation success is minimal with a single female. We, therefore, plan to synchronize the rearing of each aphid genotype to get enough “mini-colonies” every 2-3 days along the infestation experimental period. We will need at least 40 synchronized “mini-colonies” of each aphid genotype per day to perform the cross-infestation schedule (Figure 5).

For the synchronization, we will place each of the nine aphid genotypes on Golden Delicious trees grafted onto an M9 Pajam®. Note that we will be able to test for the effect of genetic proximity of the Golden Delicious cultivar used for the rearing to the other apple genotypes used for the experiment as we have sequenced the genomes of the apple genotypes used in this study. For the rearing, one Golden Delicious tree will host a given aphid genotype. After two weeks of colony growth, we will expect to have enough females to start the aphid synchronization for each genotype. Once we will get enough adult females (10-20), we will synchronize the rearing for each aphid genotype (Figure 5). The aphid synchronization aims to ensure the same developmental stage of the females and larvae that will be infested on a plant. Aphid synchronization will start mid-March

2021. Details of the synchronization procedure are described in Figure 5. For each aphid genotype (Figure 4), we will launch the aphid synchronization gradually every 2-3 days on different leaves of a Golden Delicious tree. Daily aphid synchronization is indeed challenging in such a large experimental design. We, therefore, plan to synchronize our rearing every three-four days.

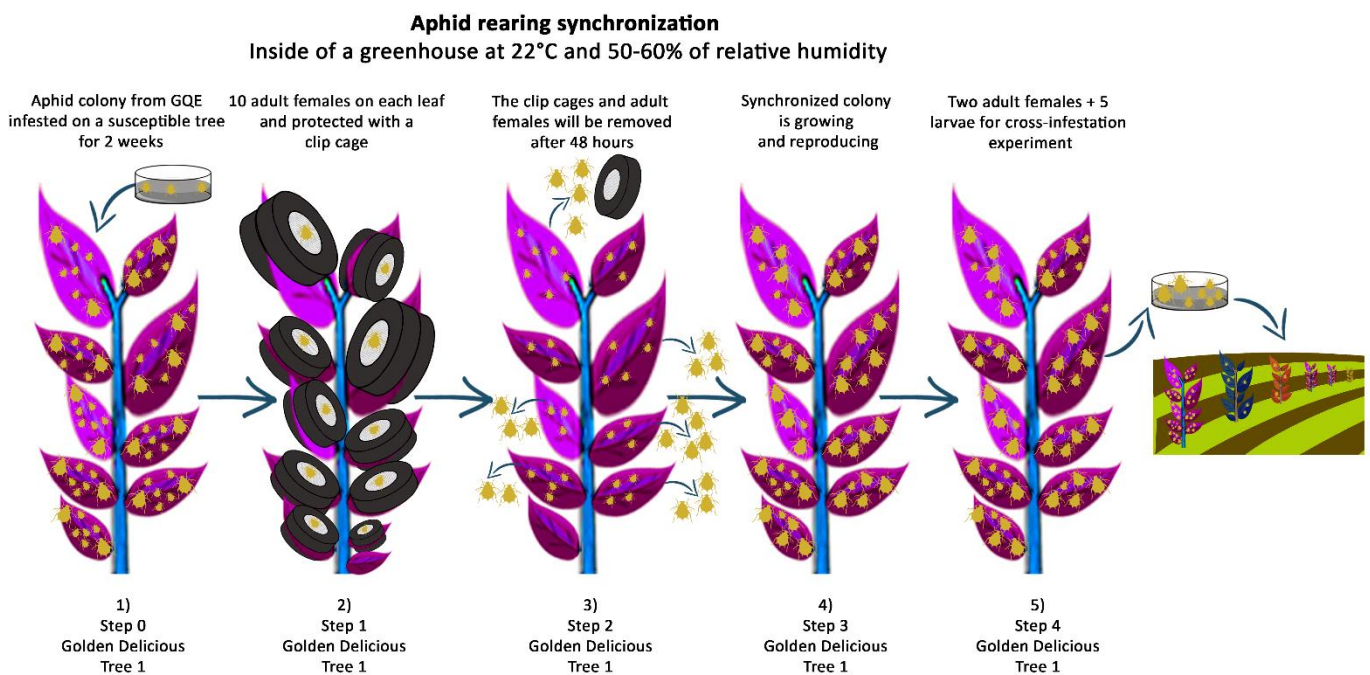
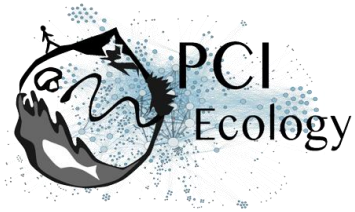


Figure 5. Aphid rearing synchronization steps explained for clone 1 from Spain. We will follow the same protocol for each aphid genotype. Step 0: a colony from the GQE-Le Moulon laboratory is received and placed onto an M9 grafted Golden Delicious susceptible apple genotype. The colony will grow for two weeks. Step 1: Ten adult females are put on a new M9 grafted Golden Delicious susceptible apple genotype, separately on different leaves, for 48 h and protected by a clip cage. Step 2: after 48 h, the clip cages and the adult females are removed and put back on the tree 1. The larvae are let grown for 10-12 days. Step 3: The larvae have grown and became adults and have started to produce larvae themselves. A synchronized colony of a single aphid genotype now grows on the tree. Step 4: Two adult females and five larvae will be selected to infest a leaf of each tree on the field. Steps 1 to 4 will be repeated every two or three days to synchronize aphid colonies for about 18 days to follow the infestation plan (see section “Aphid genotypes and preparation for the infestation”).

Detailed of modalities 1 and 2

As previously explained, we will test two modalities for each common garden (Figure 3):

-modality 1: apple genotypes infested with rosy apple aphids from different origins; seven to nine replicates of the 28 genotypes.



-modality 2: apple genotypes free of rosy apple aphid infestations; three replicates of the 28 genotypes that will be used as non-infested controls.

Modality 1: infestation, no treatment against aphids.

This modality will consist of the infestation of a mini colony (two females and five larvae) by each of the nine aphid genotypes on nine different leaves on each of the 28 apple genotypes. Each mini-colony will be isolated using a clip-cage. Note that preliminary tests in our lab show that these clip cages do not influence aphid behavior (Florencio-Ortiz et al., 2018). Each leaf will be infested with a single aphid genotype from either Belgium, France, or Spain (Figure 3). The infestation will be performed in early April 2021. Starting early April will allow us to avoid as much as possible attacks or colonization by natural enemies and other apple aphid species.

Because the aphid life cycle may vary with the climatic conditions among sites, at each site we will observe the duration of the aphid life cycle from adult to daughter-adult on a “time infestation control” cultivated apple genotype (Table 1), *i.e.*, a susceptible Golden Delicious genotype (Miñarro and Dapena, 2008). At the beginning of the cross-infestation experiment, for each of the seven to nine lines (Figure 3), a Golden Delicious apple tree will be first systematically infested with an adult female aphid. This “reference” Golden Delicious will allow us to determine what standard duration of aphid infestation will be taken for that site, *i.e.*, what will be the time to wait after an infestation to collect the colonies for each site. This duration is usually between nine to 12 days after initial infestation (Warneys et al., 2018). After this duration determined for each site, we will cut off each infested leaf together with the clip cage. Then, we will disassemble the clip cage to take the leaf with the aphid colony and transfer them into a Falcon tube previously filled with ethanol 96%. In the laboratory, we will count the number of adults and nymphs with the software ImageJ (Schneider et al., 2012).

Modality 2: control without infestation, treatment against aphids

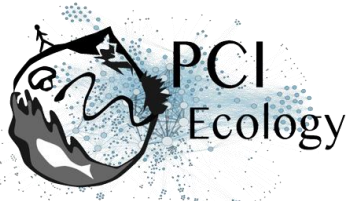
This modality will consist of the same 28 apple genotypes, not infested (Figure 3), repeated three times (Figure 3). On this modality, we will record the flowering time and bursting time.



Peer Community In Ecology

Table 2. Number of leaves infested with aphids planned in the Spring of 2021 at each common garden orchard in Belgium, France, and Spain, on each of the 28 apple genotypes (*Malus domestica* and *Malus sylvestris*, respectively). The apple genotypes included 15 *M. domestica* genotypes: five genotypes from Belgium (B1 to B5), five genotypes from France (F1 to F5), and five genotypes from Spain (S1 to S5); three tolerant apple genotypes from France (T1 to T3: two *M. domestica* apple genotypes, ‘Priscila’ cv. and ‘Florina’ cv., and one *Malus floribunda* Siebold ex Van Houtte); one susceptible genotype “Golden Delicious” (GD); Nine European wild apple genotypes *M. sylvestris* (W1 to W9, six from Belgium and three from Spain). For the aphid, three genotypes per locality, with BE_X = Belgian aphid genotype X; FR_X = French aphid genotype X; SP_X = Spanish aphid genotype X. Sympatric combinations are highlighted in grey and allopatric combinations are not highlighted.

Common garden	<i>Malus domestica</i>																		Controls				<i>Malus sylvestris</i>									Overall I									
	Belgian tres							French trees							Spanish trees				Resistant			Susceptible	Belgian trees						Spanish trees												
	Aphid	B1	B2	B3	B4	B5	SUM	Aphid	F1	F2	F3	F4	F5	SUM	Aphid	S1	S2	S3	S4	S5	SUM	Aphid	R1	R2	R3	GD1	W1	W2	W3	W4	W5	W6	W7	W8	W9	SUM	SUM				
Belgium	BE_1	9	9	9	9	8	44	BE_1	9	9	9	9	9	45	BE_1	9	9	9	8	9	44	BE_1	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	111	244	
	BE_2	9	9	9	9	8	44	BE_2	9	9	9	9	9	45	BE_2	9	9	9	8	9	44	BE_2	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	111	244	
	BE_3	9	9	9	9	8	44	BE_3	9	9	9	9	9	45	BE_3	9	9	9	8	9	44	BE_3	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	111	244	
	FR_1	9	9	9	9	8	44	FR_1	9	9	9	9	9	45	FR_1	9	9	9	8	9	44	FR_1	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	111	244	
	FR_2	9	9	9	9	8	44	FR_2	9	9	9	9	9	45	FR_2	9	9	9	8	9	44	FR_2	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	9	111	244
	FR_3	9	9	9	9	8	44	FR_3	9	9	9	9	9	45	FR_3	9	9	9	8	9	44	FR_3	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	9	111	244
	SP_1	9	9	9	9	8	44	SP_1	9	9	9	9	9	45	SP_1	9	9	9	8	9	44	SP_1	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	9	111	244
	SP_2	9	9	9	9	8	44	SP_2	9	9	9	9	9	45	SP_2	9	9	9	8	9	44	SP_2	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	9	111	244
	SP_3	9	9	9	9	8	44	SP_3	9	9	9	9	9	45	SP_3	9	9	9	8	9	44	SP_3	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	9	111	244
	SUM						396	SUM						405	SUM						396	SUM				999	2196														
Spain	BE_1	9	9	9	9	9	45	BE_1	9	9	9	9	9	45	BE_1	9	9	9	8	9	44	BE_1	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	112	246		
	BE_2	9	9	9	9	9	45	BE_2	9	9	9	9	9	45	BE_2	9	9	9	8	9	44	BE_2	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	112	246	
	BE_3	9	9	9	9	9	45	BE_3	9	9	9	9	9	45	BE_3	9	9	9	8	9	44	BE_3	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	112	246	
	FR_1	9	9	9	9	9	45	FR_1	9	9	9	9	9	45	FR_1	9	9	9	8	9	44	FR_1	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	112	246	
	FR_2	9	9	9	9	9	45	FR_2	9	9	9	9	9	45	FR_2	9	9	9	8	9	44	FR_2	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	112	246	
	FR_3	9	9	9	9	9	45	FR_3	9	9	9	9	9	45	FR_3	9	9	9	8	9	44	FR_3	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	112	246	
	SP_1	9	9	9	9	9	45	SP_1	9	9	9	9	9	45	SP_1	9	9	9	8	9	44	SP_1	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	112	246	
	SP_2	9	9	9	9	9	45	SP_2	9	9	9	9	9	45	SP_2	9	9	9	8	9	44	SP_2	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	112	246	
	SP_3	9	9	9	9	9	45	SP_3	9	9	9	9	9	45	SP_3	9	9	9	8	9	44	SP_3	8	8	9	9	9	7	9	8	8	9	9	9	9	9	9	9	112	246	
	SUM						405	SUM						405	SUM						396	SUM				1008	2214														



France	Aphid	B1	B2	B3	B4	B5	SU M	Aphid	F1	F2	F3	F4	F5	SU M	Aphid	S1	S2	S3	S4	S5	SU M	Aphid	R1	R2	R3	GD1	W1	W2	W3	W4	W5	W6	W7	W8	W9	SU M	SUM							
	BE_1	8	8	8	8	8	40	BE_1	8	8	8	8	8	40	BE_1	8	8	8	8	8	40	BE_1	8	7	8	8	7	8	8	8	8	8	8	8	8	8	8	8	102	222				
	BE_2	8	8	8	8	8	40	BE_2	8	8	8	8	8	40	BE_2	8	8	8	8	8	40	BE_2	8	7	8	8	7	8	8	8	8	8	8	8	8	8	8	8	8	102	222			
	BE_3	8	8	8	8	8	40	BE_3	8	8	8	8	8	40	BE_3	8	8	8	8	8	40	BE_3	8	7	8	8	7	8	8	8	8	8	8	8	8	8	8	8	8	102	222			
	FR_1	8	8	8	8	8	40	FR_1	8	8	8	8	8	40	FR_1	8	8	8	8	8	40	FR_1	8	7	8	8	7	8	8	8	8	8	8	8	8	8	8	8	8	102	222			
	FR_2	8	8	8	8	8	40	FR_2	8	8	8	8	8	40	FR_2	8	8	8	8	8	40	FR_2	8	7	8	8	7	8	8	8	8	8	8	8	8	8	8	8	8	102	222			
	FR_3	8	8	8	8	8	40	FR_3	8	8	8	8	8	40	FR_3	8	8	8	8	8	40	FR_3	8	7	8	8	7	8	8	8	8	8	8	8	8	8	8	8	8	102	222			
	SP_1	8	8	8	8	8	40	SP_1	8	8	8	8	8	40	SP_1	8	8	8	8	8	40	SP_1	8	7	8	8	7	8	8	8	8	8	8	8	8	8	8	8	8	102	222			
	SP_2	8	8	8	8	8	40	SP_2	8	8	8	8	8	40	SP_2	8	8	8	8	8	40	SP_2	8	7	8	8	7	8	8	8	8	8	8	8	8	8	8	8	8	102	222			
SP_3	8	8	8	8	8	40	SP_3	8	8	8	8	8	40	SP_3	8	8	8	8	8	40	SP_3	8	7	8	8	7	8	8	8	8	8	8	8	8	8	8	8	8	102	222				
	SUM						360		SUM						360		SUM						360		SUM																		918	1998

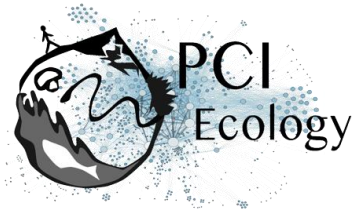
Sample size rationale

In this experiment, we have three common garden orchards located at three sites in Europe, each with five local and 10 foreign cultivated apple genotypes. Thus, we replicate local host conditions by using five independent cultivated apple genotypes from three different areas of apple cultivation. Similarly, we use three distinct aphid clone lineages from each area of origin that will be tested and selected for their genetic differences with neutral markers expected to reflect general differentiation across their genomes. This allows us to ensure that any findings consistent with local adaptation are robust. Altogether, we will have 216 sympatric combinations and 423 allopatric combinations, which provides adequate power for testing local adaptation (Kaltz and Shykoff, 1998; Kaltz et al., 1999): we will have 2/3 of allopatric comparisons (*i.e.*, aphid genotypes infested on their foreign apple genotypes and environments) against 1/3 sympatric comparisons (*i.e.*, aphid genotypes infested on their local apple genotypes and environments) (Table 2). Eventually, we will have 6,408 infestation spots (single aphid genotype on a single apple genotype leaf) in the three common gardens: 2,196 in Belgium on 244 apple trees, 1,998 in France on 222 trees, and 2,214 in Spain on 246 trees (Table 2).

We choose to perform all infestation treatments with all aphid genotypes on each apple tree. This minimizes the error variance associated with differences among trees due to their physical condition or microsite variation and therefore maximizes our power to detect differences among aphid genotypes, apple genotypes, and common garden orchards. We replicate the number of infestations as much as is logistically possible to maximize the reliability of our measures of aphid performance on a particular apple genotype at a particular site. This setup that maximizes the number of combinations, with a multi-genotype test per single tree can induce a systemic response of apple trees that can impact the fitness of a given aphid genotype within each apple tree. To take partially into account the systemic response of apple trees, each aphid genotype will be infested on the upper, middle, and lower parts of the tree. The level of infestation of each aphid genotype will be random for each apple genotype repetition. The leaf level effect will allow accounting for differential systemic apple response depending on the part of the tree infested. Note however that apple trees are infested by multiple pathogens and aphid species along the season (Miñarro et al., 2005; Alhmedi and Beliën, 2016; Tan et al., 2021), so our design is not so far from the “natural setting”. We will also perform all the infestations of each apple tree following the proposed infestation schedule. Finally, we do replicate our common garden orchards within the different areas of origin, *i.e.*, Belgium, France, and Spain. Therefore, we can adequately test the existence of local adaptation.

Manipulated variables

We manipulate the species host, the genotype of the cultivated and wild apples, the origin of the rosy apple aphids, and the sites of origin of the common garden orchards.



Apples used in this study will be of either cultivated (*M. domestica*) or wild (*M. sylvestris*) apples, with different genotypes for each. The cultivated apple genotypes were selected to represent local genotypes genetically far from each other and showing variability in the response against rosy apple aphid attacks. For the wild apple genotypes, we chose them because of the already-characterized population genetic differentiation that has been observed in the European wild apple (Cornille et al 2015). We however acknowledge that the current experiment will give a first insight into the natural response of the wild apple genotypes to the attacks of the rosy apple aphid.

We will select three different rosy apple aphid genotypes from each common garden orchard (*i.e.*, Belgium, France, and Spain) once they will be genetically characterized. To that end, we will use recently developed microsatellite markers for *D. plantaginea* to select the aphid genotypes with contrasting alleles to use for the infestation experiment.

The sites chosen for settling the common garden orchards represent a European latitudinal gradient to test the effect of local environments on the rosy apple aphid adaptation.

Measured variables

Rosy apple aphid fitness: we will measure aphid fitness for each of the nine rosy apple aphid genotypes infested on the 28 apple genotypes. The aphid fitness (W) will be calculated as follows (Warneys et al., 2018):

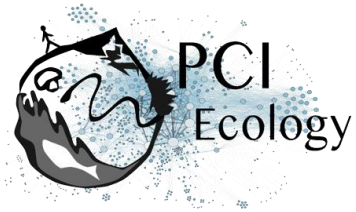
$$W = \frac{n(\text{nymphs at end of infestation}) - n_{\text{aphid}}(\text{nymphs at beginning of infestation})}{\text{total number of day of infestation}}$$

We will also, if possible, count the different insect life stages (*i.e.*, aphid larvae (L1 to L5), apterous adults, and winged forms (Angeli and Simoni, 2006)). This will be done by scaling the individuals into three categories: big (apterous females), small (larvae), winged.

Additional measurements: we will record the temperature and humidity during the experiment with a local meteorological station available next to each common garden. We will also record the temperature of each leaf, before, during, and after the infestation with Near-infrared Spectroscopy (NIRS).

Design plan and blinding

Our design included randomization. Three persons will be involved in the experiment at each common garden (Belgium, France, and Spain). Thus, people will be aware of our treatments. However, we randomized the experiment as most as possible: the infestation spot of the aphid genotype (leaf of apple genotype infested with a single aphid genotype) and the coordinates of the apple trees within each block were previously randomized. In addition, we have recorded the localization of each apple tree at each common garden orchard. Now that they are planted and growing, the initial labels attached to each tree will be removed. The trees will then have a genotype code that will not reveal the provenance or species of the apple tree during data



collection. We will control for the leaf stage and sampler effect in our statistical models, as well as the time (day and hour) of infestation.

People involved during the processing of the data will be aware of the treatments of our experiment. The design was randomized as much as possible, and the recorder effect will be tested in the statistical models, if a recorder there will be, it will be added to the equations presented in statistical model section. Moreover, the trees will have a genotype code that will not reveal the provenance or species of the apple tree during data collection. Therefore, people infesting apple trees, counting aphids, and assessing leaf damage will not know which combination is sympatric *versus* allopatric.

Study design

The study design is described in Figures 2, 3, and 4, and Tables 1 and 2.

Randomization

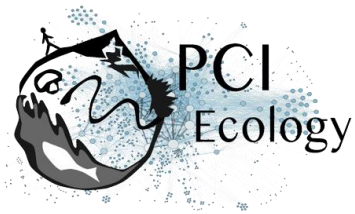
We will use replicated common gardens in three countries, Belgium, France, and Spain. Each of these experimental fields will be made of rows with randomized apple trees to prevent spatial autocorrelation of error variance from being confounded with genotypic effects. The global view of the aphid cross-infestation experiment is described in Figure 2.

Statistical models

Combining the data of the three common gardens, we will confront sympatric combinations (*i.e.*, aphid genotypes infested on apple genotypes and environments of the same origin: France, Belgium, or Spain) against allopatric combinations (*i.e.*, aphid genotypes infested on apple genotypes and environments of a different origin: France, Belgium, and Spain). We will also consider that an aphid population is locally adapted to its host and environment if its fitness is the highest on its local host and environment (Figure 1).

We will use a generalized linear mixed model (GLMM) including different factors according to the question and hypothesis that we will aim to answer. In this GLMM, the aphid genotype and apple genotype will be used as random effects, as well as the day and hour of infestation and the leaf level effect. The other effects will be fixed (see below). Then, we will gradually remove interactions and effects according to their significance. In addition, we will evaluate the differences in the effect on aphid fitness using a contrast analysis.

To test the existence of local adaptation, we will partition the three-way interaction among sites (common garden orchards), apple origin, and aphid origin into sympatric *versus* allopatric comparisons. This sympatric versus allopatric contrast will also be performed within each locality, *i.e.*, separately for the three



different common garden orchards in a similar way, in order to determine whether there is local adaptation at the different sites.

The linear mixed model that we will use to tackle each of our research questions and hypotheses are described below:

Question 1- ($G_{\text{parasite}} * \text{local environment}$): $\text{aphid_origin}_h * \text{site}_j$

Question 2- ($G_{\text{parasite}} * G_{\text{host}}$): $\text{aphid_origin}_h * \text{apple_origin}_i$

Question 3 - ($G_{\text{parasite}} * G_{\text{host}} * \text{local environment}$): $\text{aphid_origin}_h * \text{apple_origin}_i * \text{site}_j$

The following factors will be used

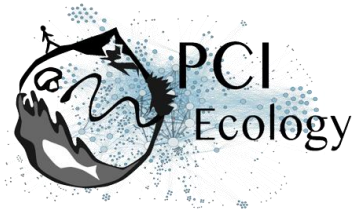
Equation 1

$W_{hijklmnot2z} = \mu_w + \text{aphid_origin}_h + \text{apple_origin}_i + \text{site}_j + \text{site}_j(\text{block}_k) + Gh_l(\text{leaf}_m(Gp_n)) + \text{day_of_infestation}_t + \text{hour_of_infestation}_{t2} + \text{leaf_level}_m + \text{tree_clone}_o + \text{aphid_origin}_h * \text{site}_j + \text{aphid_origin}_h * \text{apple_origin}_i + \text{aphid_origin}_h * \text{apple_origin}_i * \text{site}_j + \epsilon_{hijklmnot2z}$.

Mathematic equation:

$W_{hijklmnot2z} = \alpha_h + \beta_i + \gamma_j + B_{jk} + P_{lmn} + \delta_t + \zeta_{t2} + \mu_m + \omega_o + \alpha_h * \gamma_j + \alpha_h * \beta_i + \alpha_h * \beta_i * \gamma_j + \epsilon_{hijklmnot2z}$.

Table 3 describes the indexes, terms, and the effect included in our proposed GLMM. Where $W_{hijklmnot2z}$ is the absolute fitness value of an aphid genotype Gp (*i.e.* parasite genotype) from the country of origin n infested on the apple genotype l , apple tree clone o , in block k on *leaf level* m and in the common garden j infested at day t and hour $t2$, *leaf_level* $_m$ is the position of the infested leaf in the apple tree (upper, middle or lower), *tree_clone* $_o$ is the clone o of the apple genotype l , μ_w is the mean absolute fitness, *site* $_j$ is the common garden location (Belgium, Spain, France), *block* $_k$ is the block effect within each site for modality 1, *aphid_origin* $_h$ is the country of origin of the aphid (Spain, France, Belgium), *apple_origin* $_i$ is the country of origin of the apple genotype (Spain, France, Belgium), Gh_l is the apple genotype (*i.e.*, apple cultivar name) and $\epsilon_{hijklmnot2z}$ is the residual term. *Block* is random and nested within the site, and aphid *genotype* $_n$ is nested within *leaf_level* $_m$, and *leaf_level* $_m$ is nested within apple genotype Gh_l , and they were added to the models as random-effect terms. The *leaf_level* $_m$ effect is also added as a random factor alone to account for the global variability in aphid fitness that is explained by the levels at which each aphid colony was infested, whatever the apple genotypes. The *site* term measures the quality or suitability of the common garden locations, *aphid_origin* and *apple_origin* accounts for differences in fitness intrinsic to each local aphid genotype and apple genotype country of origin, *aphid_origin* $_h * \text{site}_j$ accounts for differences in local adaptation to the environment among the three aphid origins, *aphid_origin* $_h * \text{apple_origin}_i$ account for differences in local adaptation to the host among the three aphid origins, *aphid_origin* $_h * \text{apple_origin}_i * \text{site}_j$ accounts for differences in local adaptation to the host and environment among the three aphid origins. The *day_of_infestation* $_t$ and the *hour_of_infestation* $_{t2}$ consider the effect of the infestation time of the aphid genotype Gp from the country-



of-origin n on the apple genotype l in block k on leaf m and in the common garden j . We will run our proposed model using three different measures of absolute fitness (W): colony growth rate, and if possible, aphid sizes and aphid developmental stages. Note that we will measure the temperature of each apple leaf before and after aphid infestation. Temperature measured for each leaf will be first added as a fixed effect in a linear mixed model depicted in Equation 1 but without the site effect. If any effect is detected, the temperature will be added in Equation 1 as a covariance-variance matrix of a site random effect.

Question and hypothesis 4: testing in the model the $\text{aphid_origin}_h * \text{crop_wild_status}_y$ interaction.

Equation 2

$$W_{hijklmnott2yz} = \mu_W + \text{aphid_origin}_h + \text{crop_wild_status}_y + \text{site}_j + \text{site}_j(\text{block}_k) + Gh_l(\text{leaf}_m(\text{Gp}_n)) + \text{day_of_infestation}_t + \text{hour_of_infestation}_{t2} + \text{leaf_level}_m + \text{tree_clone}_o + \text{aphid_origin}_h * \text{site}_j + \text{aphid_origin}_h * \text{crop_wild_status}_y + \text{aphid_origin}_h * \text{crop_wild_status}_y * \text{site}_j + \varepsilon_{hijklmnott2yz}$$

Mathematic equation:

$$Y_{hijklmnott2yz} = \alpha_h + \eta_y + \gamma_j + B_{jk} + P_{lmn} + \delta_t + \zeta_{t2} + \mu_m + \omega_o + \alpha_h * \gamma_j + \alpha_h * \eta_y + \alpha_h * \eta_y * \gamma_j + \varepsilon_{hijklmnott2yz}$$

Question and hypothesis 5: testing in the model $\text{aphid_origin}_h * \text{tolerant_status}_x * \text{site}_j$ interaction

Equation 3

$$W_{hxijklmnott2xz} = \mu_W + \text{aphid_origin}_h + \text{tolerant_status}_x + \text{site}_j + \text{leaf_level}_m + \text{site}_j(\text{block}_k) + Gh_l(\text{leaf}_m(\text{Gp}_n)) + \text{day_of_infestation}_t + \text{hour_of_infestation}_{t2} + \text{tolerant_status}_x * \text{site}_j + \text{tree_clone}_o + \text{aphid_origin}_h * \text{tolerant_status}_x + \text{aphid_origin}_h * \text{tolerant_status}_x * \text{site}_j + \varepsilon_{hxijklmnott2xz}$$

Mathematic equation:

$$Y_{hxijklmnott2xz} = \alpha_h + K_x + \gamma_j + \omega_o + \mu_m + B_{jk} + P_{lmn} + \delta_t + \zeta_{t2} + \alpha_h * \gamma_j + \alpha_h * K_x + \alpha_h * K_x * \gamma_j + \varepsilon_{hxijklmnott2xz}$$

We will transform our future data depending on the normality and dispersion of the residuals in our models.



Peer Community In Ecology

Table 3. Description of the indexes, terms, and the effects included in our proposed statistical model General Linear Mixed Model (GLMM) to test for local adaptation of the rosy apple aphid (*Dysaphis plantaginea*).

Math index	Index	Term	Effect
α	H	Aphid_origin _h	Aphid country of origin (Spain, France, Belgium), fixed effect
β	I	Apple_origin _i	Apple country of origin (Spain, France, Belgium), fixed effect
γ	J	Site _j	Common garden site (Spain, France, Belgium), with a covariance-variance matrix of difference of temperature (or humidity) between each apple tree before (or after) the infestation, fixed effect
B	K	Block _k	Block (each block consists of 28 apple genotypes infested with 9 aphid genotypes), random effect
	L	Gh _l	Apple host genotype, random effect
μ	m	Leaf_level _m	Leaf level (Position of the infested apple leaf on the main stem. Three levels: upper, middle, or lower), random effect
P	n	Gp _n	Aphid parasite genotype, random effect
ω	o	Tree_clone _o	Apple clone of a given genotype, random effect
δ	t	Time of infestation _t	Day of infestation, random effect
ζ	t2	Time of infestation _{t2}	Hour of infestation, random effect
K	x	Tolerant_status _x	Tolerant or susceptible genotype status assessed from previous studies (Miñarro and Dapena, 2008), fixed effect
η	y	Crop_wild_status _y	Cultivated or wild apple host (<i>Malus domestica</i> and <i>Malus sylvestris</i> , respectively), fixed effect
	z		Effect of each observation
ϵ			Residual error

Inference criteria

As explained in statistical model section, we will consider multiple variables, factors, and interactions in our statistical models.

Data exclusion

We will not exclude data. We will transform our data to fit the homoscedasticity of the residuals. If there is an outlier, *e.g.*, one observation that looks vastly different from the other ones, we will first check whether there could have been any mistake. We will come back to the tubes in which each colony is conserved to count and check the number of aphids to control for mistakes. If the outlier is still valid, we will further investigate this number.

Missing data

The lack of aphids on a leaf will be a key-value, this will be counted as a true observation, *i.e.*, the absence of growth (*i.e.*, less than the original two aphid females and 5 larvae per leaf), death aphids. We will utilize Poisson, Gaussian distribution, or two-steps modeling approach with a binomial response (1 = aphid colony; 0 = absence of aphid colony) and the analysis of the aphid counting data depending on the subset of non-zero outcomes. Nevertheless, we will try to minimize recording zero in our data to avoid unnecessary data transformation. Aphids are overly sensitive to any change in environmental conditions and some infestation might fail for a technical reason. Then, if after one day of infestation the female has died, we will consider that the infestation has failed. In the case of a technical issue, we will infest again the next day and we will note this re-infestation and take it into account for statistical analyses (statistical model section). We will check if the female aphid died because of a technical issue or for a biological reason.

Acknowledgements

This work is supported by the ATIP-Inserm-CNRS program. We thank the University Paris-Saclay Short-Term Fellowship fund (to SGOV). We are grateful to Charles Eric for his assistance in the apple genotype selection. We thank the equipment and technical support of the Fablab Digiscope-LISN-Université Paris-Saclay key in the production of the aphid clip cages. We would also thank the team of the institutions that are giving us invaluable support organizing and performing the present experiment. The institutions are the GQE-Le Moulon, IRHS-INRAE-Agrocampus-Ouest-Université d'Angers in France, SERIDA in Spain, and PCFruit in Belgium. As well as interns that are involved in the experiment performance Bouliez Aurelien, Cristina de los

Reyes, Ramón Franco, Etienne Porquier, Lawrence Pelgrims, Marie Bouchaud, Quentin Marcou, Valerio Cosemans. We also strongly appreciate the time and suggestions from the Peer Community in Ecology recommenders (Eric Petit) and reviewers (Alex Stemmelén, Sharon Zytynska, and one anonymous reviewer). Version 4 of this preregistration has been peer-reviewed and in principle recommended by Peer Community In Ecology (<https://doi.org/10.24072/pci.ecology.100079>).

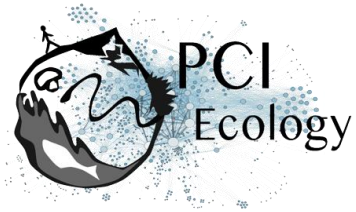
Conflict of interest disclosure

The authors of this preregistration declare that they have no financial conflict of interest with the content of this article. A. Cornille and T. Giraud are PCI Ecology recommenders.

References

- Alhmedi, A., & Beliën, T. (2016). Indirect competition in aphid community and biological control: Towards a new approach of aphid management in apple orchards.
- Angeli, G. I. N. O., & Simoni, S. (2006). Apple cultivars acceptance by *Dysaphis plantaginea* Passerini (Homoptera: Aphididae). *Journal of Pest Science*, 79(3), 175-179.
- Biello, R., Singh, A., Godfrey, C. J., Fernández, F. F., Mugford, S. T., Powell, G., ... & Mathers, T. C. (2021). A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Molecular Ecology Resources*, 21(1), 316-326.
- Bloesch, B., & Viret, O. (2013). Stades phénologiques repères des fruits à pépins (pommier et poirier). *Revue suisse de viticulture, arboriculture et horticulture*, 45(2), 128-131.
- Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic prediction of (mal) adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics*, 51, 245-269.
- Cornille, A., Gladieux, P., Smulders, M. J., Roldán-Ruiz, I., Laurens, F., Le Cam, B., ... & Gabrielyan, I. (2012). New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet*, 8(5), e1002703.
- Cornille, A., Gladieux, P., & Giraud, T. (2013). Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Applications*, 6(5), 737-748.
- Cornille, A., Giraud, T., Smulders, M. J., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, 30(2), 57-65.
- Cornille, A., Feurtey, A., Gélín, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., & Giraud, T. (2015). Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples. *Evolutionary applications*, 8(4), 373-384.

- Cornille, A., Antolín, F., Garcia, E., Vernesi, C., Fietta, A., Brinkkemper, O., ... & Roldán-Ruiz, I. (2019). A multifaceted overview of apple tree domestication. *Trends in plant science*, 24(8), 770-782.
- Dall'Agata, M., Pagliarani, G., Padmarasu, S., Troggio, M., Bianco, L., Dapena, E., ... & Tartarini, S. (2018). Identification of candidate genes at the Dp-fl locus conferring resistance against the rosy apple aphid *Dysaphis plantaginea*. *Tree genetics & genomes*, 14(1), 12.
- Fick, S. E. and Hijmans R.J., 2017. WorldClim 2: New 1-Km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12),4302-4315.
- Florencio-Ortiz, V., Novák, O., & Casas, J. L. (2018). Local and systemic hormonal responses in pepper (*Capsicum annum* L.) leaves under green peach aphid (*Myzus persicae* Sulzer) infestation. *Journal of plant physiology*, 231, 356-363.
- Hijmans, R. J., and van Etten, Jacob. 2012. raster: Geographic analysis and modeling with raster data. R package version 2.0-12. <http://CRAN.R-project.org/package=raster>
- Kaltz, O., & Shykoff, J. A. (1998). Local adaptation in host–parasite systems. *Heredity*, 81(4), 361.
- Kaltz, O., Gandon, S., Michalakis, Y., & Shykoff, J. A. (1999). Local maladaptation in the anther-smut fungus *Microbotryum violaceum* to its host plant *Silene latifolia*: evidence from a cross-inoculation experiment. *Evolution*, 53(2), 395-407.
- Koupilová, K., Koubek, T., Cornille, A., & Janovský, Z. (2021). Local maladaptation of the anther-smut fungus parasitizing *Dianthus carthusianorum*. *European Journal of Plant Pathology*, 1-10.
- Marchetti, E., Civolani, S., Leis, M., Chicca, M., Tjallingii, W. F., Pasqualini, E., & Baronio, P. (2009). Tissue location of resistance in apple to the rosy apple aphid established by electrical penetration graphs. *Bulletin of Insectology*, 62(2), 203-208.
- Miñarro, M., Hemptinne, J. L., & Dapena, E. (2005). Colonization of apple orchards by predators of *Dysaphis plantaginea*: sequential arrival, response to prey abundance and consequences for biological control. *BioControl*, 50(3), 403-414.
- Miñarro, M., & Dapena, E. (2007). Resistance of apple cultivars to *Dysaphis plantaginea* (Hemiptera: Aphididae): Role of tree phenology in infestation avoidance. *Environmental Entomology*, 36(5), 1206-1211.
- Minarro, M., & Dapena, E. (2008). Tolerance of some scab-resistant apple cultivars to the rosy apple aphid, *Dysaphis plantaginea*. *Crop Protection*, 27(3-5), 391-395.
- Olvera-Vazquez, S, G, Remoué, C., et al., (2020). Large-scale geographic survey provides insight into the colonization history of a major aphid pest on its cultivated apple host in Europe, North America and North Africa. *BioRxiv*, <https://doi.org/10.1101/2020.12.11.421644>.
- Pagliarani, G., Dapena, E., Miñarro, M., Denancé, C., et al., (2016). Fine mapping of the rosy apple aphid resistance locus Dp-fl on linkage group 8 of the apple cultivar 'Florina'. *Tree genetics & genomes*, 12(3), 56.



- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9(7), 671-675.
- Simon, J. C., d'Alencon, E., Guy, E., Jacquin-Joly, E., Jaquier, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in functional genomics*, 14(6), 413-423.
- Smadja, C. M., Canbäck, B., Vitalis, R., Gautier, M., Ferrari, J., Zhou, J. J., & Butlin, R. K. (2012). Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution: International Journal of Organic Evolution*, 66(9), 2723-2738.
- Tan, X. M., Yang, Z. S., Zhou, H., Yang, Q. M., & Zhou, H. X. (2021). Resistance performance of four principal apple cultivars to woolly apple aphid, *Eriosoma lanigerum* (Hemiptera: Pemphigidae), by simulated seasonal temperature in northern China. *Arthropod-Plant Interactions*, 15(1), 59-69.
- Warneys, R., Gaucher, M., Robert, P., Aligon, S., Anton, S., Aubourg, S., ... & Heintz, C. (2018). Acibenzolar-S-Methyl Reprograms Apple Transcriptome Toward Resistance to Rosy Apple Aphid. *Frontiers in plant science*, 9, 17

8. GENERAL DISCUSSION AND CONCLUSION

In my Ph.D. project, I used the rosy apple aphid *Dysaphis plantaginea* as a model to unravel the role of selection, drift, and gene flow during the divergence of species and populations of aphids. In particular, I evaluated the genomic and ecological bases of the rosy apple aphid adaptation (or adaptive divergence) using different approaches. First, using a comparative genomics lens, I reconstructed the evolution of major genes involved in the aphid adaptation to their host. I included the rosy apple, 11 other aphids, and one aphid-like species that allowed me to understand the long-term evolution of aphids (Chapter 1). Using a microevolutionary lens, I then investigated patterns of local adaptation in aphids at the genomic and phenotypic levels using the rosy apple aphid model. As a preliminary step, I reconstructed the colonization history of the rosy apple aphid using different genetic markers (microsatellites and SNPs, Chapters 2 and 3). This step is mandatory to avoid the confounding effect of demography and selection. Then, we sought genomic regions under selection in different rosy apple aphid populations spread across its distribution to investigate the genomic basis of local adaptation, if any (Chapter 3). Last, we tested the ecological bases of local adaptation of rosy apple aphid populations from different origins and infested on different apple cultivars in different common garden orchards in Europe (Chapter 4). The results of each chapter presented in this thesis work gave us critical information to understand the evolutionary processes involved in the (local adaptive) divergence of the rosy apple aphid and its sister species.

CHAPTER 1: CHEMOSENSORY GENES AND TRANSPOSABLE ELEMENTS EVOLUTIONARY DYNAMICS IN APHIDS

We first analyzed studies of the evolution of key genes related to the host adaptation to their hosts. We focused on two major families of chemosensory receptor genes, the gustatory receptors (or GR) and olfactory receptors (OR). These genes are the insect chemosensation system that recognizes chemicals to recognize their host or pheromones to locate and identify mates or escape their predators (Robertson et al., 2019). We selected 12 aphid species with different ecological and host preferences. We included the new genome assembly of our model species, the rosy apple aphid, which is available in the team. We also added one aphid-like relative species used as an outgroup. We showed a higher evolutionary rate in the Aphidinae subfamily driven by gene expansion (e.g., immune response, transposase activity, retrotransposon nucleocapsid) and contraction (e.g., regulation of transcription by RNA polymerase) with functions that are potentially related to the adaptation to their hosts. We observed that the evolution of GR and OR genes was in tandem following the birth-death model. We found that GR and OR genes evolved mainly under purifying selection. Only a fifth of OR and GR genes evolved under positive selection, including gustatory receptor gene groups (GR02) related to fructose detection. A negative correlation between rates of evolution of GR/OR genes and species divergence time indicated that recent positive selection is observed in recently diverged sequences. Our results suggest that periods during which OR and GR genes evolve under positive selection may occur episodically and be followed by prolonged episodes of purifying selection. Our results suggest that TEs did not have a role in fueling the diversity of the GR and OR genes. Besides, we found species-specific TE dynamics, with major TE bursts in aphids infesting fruit trees (*Dysaphis plantaginea* and *Myzus persicae*).

The aphid superfamily comprises more than 4,000 species (Gautam et al., 1993), including the subfamily Aphidinae, the most species-rich subfamily, with about 2,500 species (Li et al., 2022). Several hypotheses have been proposed to explain the high diversification in Aphidinae (Kim et al., 2011). For instance, the life cycle was presented as a driver of aphid speciation. The heteroecious state or host-alternation in some aphid groups may lead to colonizing new hosts and favoring aphid speciation. Host alternation was also thought to avoid specialization (Moran, 1987). Our analysis included species with heteroecious life cycles (e.g., *D. plantaginea*, Introduction, Figure 1) and monoecious life cycles, such as the model species *A. pisum* (Brisson and Stern, 2006). However, we did not reveal higher evolutionary rates in heteroecious aphids. We found a high evolutionary rate in a subfamily known for the variability in the life cycles and aphid species with large host breath, suggesting that host breath may have driven speciation.

We observed that OR and GR evolved mainly by purifying selection. The evolutionary timescale of our study could explain this result. Beyond the manual annotation, the most significant technical challenge I faced for the comparative genomics analyses was the synteny and high divergence among GR and OR genes. Indeed, GR and OR genes were poorly conserved among aphid species. Because of the high divergence of the GR and OR sequences between aphid genera, performing selection tests, which require aligning sequences, was challenging. The precise estimates of divergence time among aphid species are still poorly documented because of the lack of fossil aphid species (Munro et al., 2011). However, the time tree calibrated with the few fossils available revealed that the evolutionary timescale at which we focused was very large (137 – 15 Mya), and we probably missed the recent evolutionary process involved in the evolution of OR and GR genes. This can explain the lack of observations of signatures of recent positive selection in those genes. Further analyses could focus on the genus *Dysaphis* and

include closely related species found on fruit trees or perennials (e.g., *Dysaphis*, *Aphis*, *Bracchycaudus*) to investigate the evolution of OR and GR genes at a shorter evolutionary timescale than the one we looked at. For instance, studies at the genus level using comparative genomics in the tsetse flies *Glossina* spp. detected the expansion and contraction of genes with functions related to protease, odorant-binding, and helicase activities (Attardo et al., 2019). The ideal timescale to study parasite adaptation to their host is complex. Parasite (local) adaptation can result from the coevolution between the aphids and their host plants. Coevolution is a dynamic process and looking at the microevolutionary scale (at the population level) is an avenue to understand better aphid adaptation to their hosts (Cornille et al., 2022). This is why I then focused on population-level analyses during my PhD.

We also compared the TE content and evolutionary dynamics among the analyzed species. The activity and accumulation of TEs can have multiple effects on genomes, for instance, the gene disruption by TE insertions producing spontaneous mutations expressed in different phenotypes of *D. melanogaster* (Erickbush and Furano, 2022). We investigated the potential role of TEs in the evolution of chemosensory genes. However, we did not find any enrichment of TE surrounding the GR or OR genes. However, we observed that the few TE enrichments were close to OR genes. OR genes are more recent than GR genes. This suggests that our lack of observation of TE enrichment can result from purifying selection, which eliminated TEs once they fueled diversity of OR and GR genes. This hypothesis could be tested at a smaller evolutionary scale within the *Dysaphis* genus.

We proposed a manually annotated chemosensory gene set that would be key for developing sustainable aphid control management programs. However, this set of aphid chemosensory genes should be further functionally validated. A

first step can be transcriptomic analyses in aphids to study the expression of the candidate GR and OR genes. For instance, candidate chemosensory genes (GRs, ORs, IRs, OBPs) were detected using transcriptome analysis in the crop pest, the beet webworm *Loxostege stictivalis* L, and key in foraging, seeking mates, and host recognition (Wei et al., 2017). Functional characterization in mutants can also be the next step (Li et al., 2023). In addition, further studies could focus on the manual annotation of other gene families that are also part of the chemosensation system (e.g., ionotropic receptors or IRs and odorant-binding proteins or OBPs) and may also be crucial in aphid adaptation to their host. The OBPs are highly conserved in the hemipteran group. Studies of OBP genes in the brown planthopper or *Nilaparvata lugens*, a major pest of rice, showed that these genes are highly expressed in the antennae and propose an essential function in signal-binding genes for plant host recognition (Xue et al., 2014). The number of chemosensory genes, including IRs genes, was positively correlated to the host range in three thrips species (*Thripidae* Stephens spp), suggesting their essential function for detecting their host (Hu et al., 2023). The pipeline we used (Meslin et al., 2022) provides unprecedented annotations of OR and GR genes in aphids and also opens avenues to study additional candidate genes involved in aphids, such as OBP or IR.

This study provides insight into the evolutionary processes shaping the molecular evolution of genes involved in phytophagous insect-plant interactions. It also offers an unprecedented database for studying aphid genome evolution and aphid adaptation to their host. Thus, further work would increase the number of species at a smaller evolutionary scale and use chromosomal-level genome assembly to obtain a detailed view of the genomic architecture and evolution of GR and OR, together with the role of TE in fueling their evolution.

CHAPTER 2: COLONIZATION HISTORY OF THE ROSY APPLE APHID

We analyzed the microevolutionary history of the rosy apple aphid. We wanted to know if the rosy apple aphid followed the domestication history of the cultivated apple. The cultivated apple originated from Central Asia, following the Silk Route, and was brought to Europe by the Greeks around 1,500 years ago (Cornille et al., 2014, 2019). We included rosy apple aphid populations from different countries where the rosy apple aphid was reported to infest the cultivated apple (*i.e.*, Europe, North Africa, and North America). We detected five genetic groups: American, Moroccan, Spanish, Eastern European (including Italy), and Western European (including Denmark). We found rapid and recent colonization of the rosy apple aphid in Europe with gene flow between and within populations with a weak genetic spatial structure across Europe. We also observed that the rosy apple aphid rarely hosts an endosymbiont different than the primary *Buchnera aphidicola*. These findings were our basis for continuing our microevolutionary evolution analysis, seeking genomic patterns of adaptation in the rosy apple aphid.

To further study the colonization history of *Dysaphis plantaginea*, we needed to add additional samples and species of *Dysaphis* ssp. The challenge to obtaining *D. plantaginea* samples is because of the similarities in morphology and taxonomic difficulties (Heie, 1986; Blackman and Estop, 2000). *Dysaphis plantaginea* has been reported in Central Asia (CAB, 2014), possibly due to a host shift, but we failed to obtain samples from this region. However, we got additional samples from the Middle East, which is supposedly the origin of *D. plantaginea*. *Dysaphis plantaginea* may also have originated from a host jump from other hosts or populations. There are different *D. plantaginea* sister species, such as *Dysaphis radicola* Mordilko, *D. devecta*, *Dysaphis brancoi* Börner, *Dysaphis anthrisci* Börner, *Dysaphis chaerophylli*

Börner that feed on the cultivated apple, as their primary host (Blommers et al., 2004; Stekolshchikov 2006). But other aphid sister species feed on the cultivated pears, for instance, *Dysaphis reaumuri* Mordvilko and *Dysaphis pyri* Boyer de Fonscolombe (Barbagallo et al., 2007). The *Pyrus* genus is known to have diverged a long time ago from the genus *Malus*, probably in the Caucasus (Celton et al., 2009; Xiang et al., 2017). Therefore, the ancestral group from which *D. plantaginea* diverged might have had a *Pyrus* species as a host. With that in mind, we used SNPs and included additional samples of *D. plantaginea* and its sister species, *D. pyri*, and *D. devectora*, and provided further insights into the population structure of the RAA and its demographic history and origin.

Additional analyses that can also be envisioned include investigating the colonization history of the RAA by looking at TE content. From Chapter 1, we saw that *D. plantaginea* showed a burst of transposable elements. Colonizing new environments could drive recent TE activities (Kofler et al., 2015). An analysis of a major fungal wheat pathogen, *Z. tritici* (Roberge ex Desm.) Quaedvl. & Crous showed that demography history played a fundamental role in shaping the genome within *Z. tritici* populations, finding that TE insertions and genome size increased with colonization bottlenecks (Oggenfus et al., 2021). As a first step, a study of the TE copy numbers in different populations of *D. plantaginea* could provide first insights into their role in the colonization history of *D. plantaginea*. We could also study the distribution of transposon insertion (TIPs) along the genomes and confront it to signatures of selection I found in Chapter 3 to understand the adaptive role of TE in aphid colonization.

Our inferences and the perspectives discussed above will have important implications for establishing management strategies for pest control programs (e.g., treatment application) and apple breeding programs (Guillemaud et al.,

2011; Peace et al., 2019). Moreover, Chapter 2 brings information to understand the colonization history of crop pests in the frame of the domestication of fruit trees and develop strategies to deal with current and future pest outbreaks.

CHAPTER 3: GENOMIC BASES OF LOCAL ADAPTATION IN THE ROSY APPLE APHID (*DYSAPHIS PLANTAGINEA*) TO ITS HOST, THE CULTIVATED APPLE (*MALUS DOMESTICA*)

In Chapter 3, we investigated the genomic signatures of adaptive divergence among closely related species of *Dysaphis* ssp. and local adaptive divergence in *D. plantaginea*. We selected a subset of individuals representing each genetic group detected in Chapter 2, adding samples from Canada, Sweden, and Israel. We also included individuals from two sister species, *Dysaphis devectora* (sampled in Romania and Iran) and *Dysaphis pyri* (sampled in France, Spain, and Iran). We seek signatures of selection in the genomes of closely relative species with different biology, ecology, or host preferences. For instance, *D. devectora* has been recorded to infest the same host genus as *D. plantaginea*, the *Malus* genus, but *D. devectora* lacks host alternance. These species also present different biology. *Dysaphis devectora* develops a colony protected by a leaf pseudo gall produced by rolling the leaf laterally and turning it red. On the other hand, *D. pyri* has a different host and seasonal host alternate with the primary host, the *Pyrus* genus, and the secondary host, *Galium* species (Blackman and Eastop, 2000). Using SNPs, we detected a clear genetic structure among the three aphid species: *D. devectora*, *D. plantaginea*, and *D. pyri*. Within *D. plantaginea*, we showed clear genetic structure between North American, Israeli, and Moroccan/Spanish samples but weaker differentiation among samples from European countries (excluding Spain). We detected a recent population expansion and split among the *D. plantaginea* genetic groups. We also found that *D. plantaginea* split recently from *D. pyri* and *D. devectora*. We distinguished gene flow between the *D. plantaginea* genetic group from Morocco and Spain and Europe, and America and Morocco and Spain. We detected

signals of positive selection in the genomes of the six populations using two different statistics. However, we must now confront SNP outliers detected with the two statistics to validate our results and further dig into their functions. We will also assess other genomic patterns of adaptation (such as balancing selection and soft sweep).

For Chapter 3, I used an updated version of the RAA genome assembly (v4) compared to the ones used in Chapter 1 (v3). Getting a chromosomal-level assembly of the rosy apple aphid genome (Version 4) was challenging. The genome assembly of the rosy apple aphid implied the utilization of different sequencing technologies, and several research groups were involved. The rosy apple aphid genome assembly started in 2018 before I arrived at the lab. Thus, I was not directly involved in the process of the different versions of the rosy apple aphid genome assembly (aphid rearing, DNA extraction, and bioinformatic approaches) processes before the genome assembly. However, I had the opportunity to discuss the assembly updates used to process the SNP mapping and calling and learn about the filtering steps to obtain the variant call file (VCF). Assembly of insect genomes has been reported to be a challenge (Zaho et al., 2019), which is explained by gene duplications in aphids. In aphids, gene duplications are pervasive events throughout their evolution (International Aphid Genomics Consortium, 2010; Mathers et al., 2021; Julca et al., 2019). These duplications date back to the ancestral clades of aphids like in phylloxera (*D. vitifoliae*, included in our analysis), presenting high rates of old and recent duplication events (Julca et al., 2019; Rispe et al., 2020). The characteristics of gene duplications are related to gene novelties and adaptive radiation (Innan and Kondrashov, 2017). We found these duplications in the rosy apple aphid genome, which may have hampered the species assembly. However, thanks to the different teams involved in assembling the different genome versions, the chromosome-level assembly version 4 was available for

chapter 3 of my PhD thesis. This chromosome-level assembly was key to studying the genomic architecture of adaptation in *Dysaphis ssp.* Our future findings from Chapter 3 will bring information to elucidate the abiotic and abiotic patterns involved in the adaptation and speciation of important crop pests.

CHAPTER 4: EXPERIMENTAL TEST FOR LOCAL ADAPTATION OF THE ROSY APPLE APHID (*DYSAPHIS PLANTAGINEA*) DURING ITS RECENT RAPID COLONIZATION ON ITS CULTIVATED APPLE HOST (*MALUS DOMESTICA*) IN EUROPE

Eventually, we assessed the ecological bases of local adaptation of the rosy apple aphid through an experimental approach presented in Chapter 4. We used a common garden experiment in three countries with different environments. We used apple genotypes with local usage, wild relatives, and genotypes reported to be susceptible or tolerant against aphid infestation that could bring us key information to manage a major apple tree pest. We defined the aphid development stage, colony growth, and aphid size as fitness measurements of stress (Jedličková et al., 2015; Vilcinskas, 2016) to disentangle patterns of local adaptation. Based on these traits, we found a pattern of maladaptation in the French genotypes in our preliminary results; however, we are still exploring the data, adding and removing variables to confirm these findings. We expect to find adaptation patterns at the phenotypic that we could complement with the genomic findings to get support for the role of abiotic or/and biotic patterns of adaptation in the rosy apple aphid. This experiment allowed me to participate in leading a complex experimental field in Spain. As a group, we fixed problems in each common garden. After exchanging from the three countries, we dealt with the local climate with the challenge of rearing the rosy apple aphid. The rosy apple aphid prefers young shoots but is also susceptible to temperature and humidity levels. We are expecting that the results of Chapter 4 would set the basis to extend a common garden experiment with key places and key genotypes (detected from Chapters 2 and 3) to perform a

common garden cross infestation, including sister species, to elucidate the role of abiotic and biotic environment. Based on our current experience, it would be possible to understand these questions in a gradient limited to a single country but extending the number of infestation spots and aphid species.

Our work emphasizes the significance of considering both the local environment and the host when studying phytophagous insects. The information on complex interactions between apple tree genotypes, aphid parasites, and their respective environments, along with the data from parasite adaptation to its host (*i.e.*, chemoreceptors, Chapter 1), colonization history (Chapter 2), and genomic signature of local adaptation (Chapter 3) are a basis for the development of sustainable strategies against the control of a major pest of the cultivated apple tree.

8.1 GENERAL CONCLUSION

We generate key information to understand the ecological and genomic drivers of local adaptation in the plant-herbivore system. This work represents not only an academic exercise but understanding the environmental and evolutionary processes that have shaped species distribution and genetic structure are also essential to predict how organisms will face current ecological crises (*e.g.*, climate change, ecosystem disturbance, pest outbreaks; Gienapp et al., 2007; Merilä et al., 2014). We show that comparative genomics, population genomics, and ecological approaches are robust approaches to getting information about the crop evolutionary history, genetic structure, and patterns of pest adaptation to their environment. The findings of the present work are important to mitigate current and future pest outbreaks. The present work provided some answers but new questions on the evolutionary processes behind parasite (adaptive) divergence from their host. For instance, would we observe increased speciation rates associated

with the life cycles, *e.g.*, heteroecious against monoecious life cycles, if we included more aphid species presenting various life cycles? Which other genes than OR and GR are key in the adaptation of a parasite to its host (*i.e.*, IRs and OBPs)? How did OR, GR, and other candidate genes evolve at the genus level, *e.g.*, within the *Dysaphis* genus? Have the transposable elements fuelled chemosensory gene diversity at this microevolutionary timescale (compared to the large evolutionary scale studied in Chapter 1)? Is there any genomic pattern of adaptation to the local abiotic or biotic environment in sister species of *D. plantaginea*, for instance, *D. devecta* and *D. pyri*? Is *D. plantaginea* locally adapted to its local secondary host? I envision these questions as a basis to continue exploring the genomic and ecological bases of aphid adaptation.

8.2 REFERENCES

Attardo, G. M., Abd-Alla, A. M., Acosta-Serrano, A., Allen, J. E., Bateta, R., Benoit, J. B., ... & Aksoy, S. (2019). Comparative genomic analysis of six *Glossina* genomes, vectors of African trypanosomes. *Genome biology*, 20(1), 1-31. <https://doi.org/10.1186/s13059-019-1768-2>

Barbagallo, S., Cocuzza, G., Cravedi, P., & Komazaki, S. (2007). IPM case studies: deciduous fruit trees. In *Aphids as crop pests* (pp. 651-661). Wallingford UK: CABI. <https://doi.org/10.1079/9780851998190.0651>

Blackman, R. L., & Eastop, V. F. (2000). *Aphids on the world's crops: an identification and information guide* (No. Ed. 2). John Wiley & Sons Ltd.

Blommers, L. H. M., Helsen, H. H. M., & Vaal, F. W. N. M. (2004). Life history data of the rosy apple aphid *Dysaphis plantaginea* (Pass.)(Homopt., Aphididae) on plantain and as migrant to apple. *Journal of Pest Science*, 77, 155-163. <https://doi.org/10.1007/s10340-004-0046-5>

Bonnemaison L (1959) Le puceron cendré du pommier (*Dysaphis plantaginea* Pass.) Morphologie et biologie–Méthodes de lutte. *Annales de l'Institut National de la Recherche Agronomique, Série C, Epiphyties*, 3, 257–322.

Brisson, J. A., & Stern, D. L. (2006). The pea aphid, *Acyrtosiphon pisum*: an emerging genomic model system for ecological, developmental and evolutionary studies. *Bioessays*, 28(7), 747-755. <https://doi.org/10.1002/bies.20436>

Amandine, C., Ebert, D., Stukenbrock, E., de la Vega, R. C. R., Tiffin, P., Croll, D., & Tellier, A. (2022). Unraveling coevolutionary dynamics using ecological genomics. *Trends in Genetics*. <https://doi.org/10.1016/j.tig.2022.05.008>

CAB, UK, (2014). *Dysaphis plantaginea* (Passerini (1860)), rosy apple aphid.[pest/pathogen]. *Dysaphis plantaginea* (Passerini (1860)), rosy apple aphid.[pest/pathogen]., (AQB CPC record). <http://www.cabi.org/cpc/datasheet/20214>

Cornille A, Giraud T, Smulders MJ, Roldán-Ruiz I, Gladieux P (2014) The domestication and evolutionary ecology of apples. *Trends in Genetics*, 30, 57–65. <https://doi.org/10.1016/j.tig.2013.10.002>

Cornille A, Antolín F, Garcia E, Vernesi C, Fietta A, Brinkkemper O, Kirleis W, Schlumbaum A, Roldán-Ruiz I (2019) A Multifaceted Overview of Apple Tree Domestication. *Trends in Plant Science*, 24, 770–782. <https://doi.org/10.1016/j.tplants.2019.05.007>

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular ecology*, 23(13), 3133-3157. <https://doi.org/10.1111/mec.12796>

Dohlen, C. D. V., & Moran, N. A. (2000). Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biological Journal of the Linnean Society*, 71(4), 689-717. <https://doi.org/10.1111/j.1095-8312.2000.tb01286.x>

Gautam, D. C., Crema, R., & Pagliai, A. M. B. (1993). Cytogenetic mechanisms in Aphids. *Italian Journal of Zoology*, 60(3), 233-244. <https://doi.org/10.1080/11250009309355818>

Heie, O. E. (1986). *The Aphidoidea (Hemiptera) of Fennoscandia and Denmark*. Brill.

Hu, Q. L., Ye, Z. X., Zhuo, J. C., Li, J. M., & Zhang, C. X. (2023). A chromosome-level genome assembly of *Stenchaetothrips biformis* and comparative genomic analysis highlights distinct host adaptations among thrips. *Communications Biology*, 6(1), 813. <https://doi.org/10.1038/s42003-023-05187-1>

Innan, H., & Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics*, 11(2), 97-108. <https://doi.org/10.1038/nrg2689>

International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS biology*, 8(2), e1000313. <https://doi.org/10.1371/journal.pbio.1000313>

Jedlička, P., Jedličková, V., & Lee, H. J. (2015). Expression of stress-related genes in the parthenogenetic forms of the pea aphid, *Acyrtosiphon pisum*. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 180, 32-37. <https://doi.org/10.1016/j.cbpa.2014.11.009>

Kim, Y., & Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics*, 167(3), 1513-1524.

Kim, H., Lee, S., & Jang, Y. (2011). Macroevolutionary patterns in the Aphidini aphids (Hemiptera: Aphididae): diversification, host association, and biogeographic origins. *PLoS one*, 6(9), e24749. <https://doi.org/10.1371/journal.pone.0024749>

Kofler, R., Nolte, V., & Schlötterer, C. (2015). Tempo and mode of transposable element activity in *Drosophila*. *PLoS Genetics*, 11(7), e1005406. <https://doi.org/10.1371/journal.pgen.1005406>

Lin, R., Yang, M., & Yao, B. (2022). The phylogenetic and evolutionary analyses of detoxification gene families in Aphidinae species. *Plos one*, 17(2), e0263462. <https://doi.org/10.1371/journal.pone.0263462>

Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., ... & Walters, J. R. (2019). Insect genomes: progress and challenges. *Insect molecular biology*, 28(6), 739-758. <https://doi.org/10.1111/imb.12599>

Li, Z., Capoduro, R., Bastin-Héline, L., Zhang, S., Sun, D., Lucas, P., ... & Meslin, C. (2023). A tale of two copies: Evolutionary trajectories of moth pheromone receptors. *Proceedings of the National Academy of Sciences*, 120(20), e2221166120. <https://doi.org/10.1073/pnas.2221166120>

Mathers, T. C., Wouters, R. H., Mugford, S. T., Swarbreck, D., Van Oosterhout, C., & Hogenhout, S. A. (2021). Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Molecular Biology and Evolution*, 38(3), 856-875. <https://doi.org/10.1093/molbev/msaa246>

Meslin, Camille, Pauline Mainet, Nicolas Montagné, Stéphanie Robin, Fabrice Legeai, Anthony Bretaudeau, J Spencer Johnston, et al. 2022. "Spodoptera Littoralis Genome Mining Brings Insights on the Dynamic of Expansion of Gustatory Receptors in Polyphagous Noctuidae." *G3 Genes|Genomes|Genetics*. <https://doi.org/10.1093/g3journal/jkac131>

Moran, N. A. (1988). The evolution of host-plant alternation in aphids: evidence for specialization as a dead end. *The American Naturalist*, 132(5), 681-706. <https://doi.org/10.1086/284882>

Munro, J. B., Heraty, J. M., Burks, R. A., Hawks, D., Mottern, J., Cruaud, A., ... & Jansta, P. (2011). A molecular phylogeny of the Chalcidoidea (Hymenoptera). *Plos one*, 6(11), e27023. <https://doi.org/10.1371/journal.pone.0027023>

Oggenfuss, U., Badet, T., Wicker, T., Hartmann, F. E., Singh, N. K., Abraham, L., ... & Croll, D. (2021). A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *Elife*, 10, e69249. <https://doi.org/10.7554/eLife.69249>

Rispe, C., Legeai, F., Nabity, P. D., Fernández, R., Arora, A. K., Baa-Puyoulet, P., ... & Tagu, D. (2020). The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC biology*, 18(1), 1-25. <https://doi.org/10.1186/s12915-020-00820-5>

Robertson, H. M., Robertson, E. C., Walden, K. K., Enders, L. S., & Miller, N. J. (2019). The chemoreceptors and odorant binding proteins of the soybean and pea aphids. *Insect biochemistry and molecular biology*, 105, 69-78. <https://doi.org/10.1016/j.ibmb.2019.01.005>

Vilcinskis, A. (Ed.). (2016). *Biology and ecology of aphids*. CRC Press.

Wei, H. S., Li, K. B., Zhang, S., Cao, Y. Z., & Yin, J. (2017). Identification of candidate chemosensory genes by transcriptome analysis in *Loxostege sticticalis* Linnaeus. *Plos one*, 12(4), e0174036. <https://doi.org/10.1371/journal.pone.0174036>

Xue, J., Zhou, X., Zhang, C. X., Yu, L. L., Fan, H. W., Wang, Z., ... & Cheng, J. A. (2014). Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. *Genome Biology*, 15(12), 1-20. <https://doi.org/10.1186/s13059-014-0521-0>