



HAL
open science

Genetic bases and evolutionary consequences of sex loss in aphids

Hélène Defendini

► **To cite this version:**

Hélène Defendini. Genetic bases and evolutionary consequences of sex loss in aphids. Agricultural sciences. Agrocampus Ouest, 2023. English. NNT : 2023NSARA094 . tel-04617107

HAL Id: tel-04617107

<https://theses.hal.science/tel-04617107v1>

Submitted on 19 Jun 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.

THESE DE DOCTORAT DE

L'INSTITUT AGRO RENNES ANGERS

ECOLE DOCTORALE N° 600

Ecologie, Géosciences, Agronomie, Alimentation

Spécialité : *Écologie et évolution*

Par

Hélène DEFENDINI

Bases génétiques et conséquences évolutives de la perte de sexe dans le groupe des pucerons

Thèse présentée et soutenue à Rennes, le 08/12/2023

Unité de recherche : UMR IGEPP, Le Rheu

Thèse N° : A-94 2023-25

Composition du Jury :

Rapporteurs avant soutenance :

Karine Van Doninck

Professeure, Université libre de Bruxelles

Denis Roze

Chargé de recherche, CNRS - Station Biologique de Roscoff

Examineurs :

Thomas Lenormand

Directeur de recherche, CNRS - CEFE Montpellier

Christelle Fraïsse

Chargée de recherche, CNRS - Université de Lille

Manuel Plantegenest

Professeur, INRAE - IGEPP, Institut Agro Rennes

Directeur de thèse :

Jean-Christophe Simon

Directeur de recherche, INRAE - IGEPP, Le Rheu

Co-directrice de thèse :

Julie Jaquiéry

Chargée de recherche, INRAE - IGEPP, Le Rheu

Table of contents

Table of contents	1
Acknowledgments / Remerciements	3
Introduction	4
1. Genetic basis of sex loss in eukaryotes	5
2. Evolutionary consequences of sex loss in eukaryotes	7
2.1. Effect of the lack of genomic recombination in clonal asexual species	7
2.1.1. Genome evolution in non-recombining asexual species	7
2.1.2. Gene expression evolution in non-recombining asexual species	8
2.2. Effect of the absence of sexual morphs	9
2.2.1. Gene sequence evolution in asexuals.....	10
2.2.2. Gene expression evolution in asexuals.....	10
3. Aphids as models for evolutionary studies of asexuality	11
4. Thesis objectives	12
5. References	14
Chapter 1 – Contrasting evolutionary patterns between sexual and asexual lineages in a genomic region linked to reproductive mode variation in the pea aphid	19
Abstract	20
Introduction	21
Results	22
Discussion	28
Materials and methods	32
References	35
Supplementary materials	38
Chapter 2 – To each his own: no evidence of shared genetic basis for sex loss among different aphid species	51
Abstract	52
Introduction	53
Material and methods	55
Results	60
Discussion	69
References	72
Supplementary materials.....	79

Chapter 3 – Evolutionary consequences of loss of sexual reproduction on male-related traits in parthenogenetic lineages of the pea aphid	81
Abstract	82
Introduction	83
Materials and methods	84
Results	87
Discussion	90
References	93
Supplementary materials	96
Chapter 4 – The release of sexual conflict after sex loss is associated with evolutionary changes in gene expression	99
Abstract	100
Introduction	101
Materials and methods	103
Results	106
Discussion	111
References	115
Supplementary materials	119
Discussion	133
1. Genetic bases of sex loss	133
1.1. Multiple and independent losses of sexual reproduction in aphids	133
1.2. Asexual lineages retain evolutionary potential	134
1.3. Perspectives	135
2. Evolutionary consequences of sex loss	135
2.1. Transcriptomic and phenotypic changes in asexual lineages are associated with sex loss	136
2.2. Deciphering the nature and timeline of evolutionary forces underlying these changes	136
2.3. Perspectives.....	138
3. Conclusion	139
4. References	140
Annex	143
Annex 1. Congress poster - <i>Evolution of gene expression following sex loss in the pea aphid</i>	143
Abstract.....	145
Résumé.....	146

Remerciements

Je remercie en premier lieu mon directeur et ma co-directrice de thèse, Jean-Christophe et Julie, qui m'ont fait découvrir de la meilleure façon possible le monde de la recherche. Merci pour votre grande disponibilité et la confiance accordée à mon travail.

Je tiens également à remercier l'ensemble des membres du jury d'avoir accepté d'évaluer mes travaux de thèse, Karine Van Doninck et Denis Roze en tant que rapporteur·rice-s, ainsi que Christelle Fraïsse, Thomas Lenormand et Manuel Plantegenest en tant qu'examineur·rice-s.

Les membres de mon comité de thèse, Sylvain Glémin, Tanja Schwander, Thomas Broquet et Christoph Haag, m'ont donné de précieux conseils à plusieurs reprises au cours de ma thèse, je les remercie chaleureusement.

I am also very grateful to Chris Bass for his valuable contribution to my PhD, both in terms of the data he provided and his helpful comments on my results.

Merci à toute l'équipe EGI pour son aide dans les expériences et/ou analyses, notamment Fabrice L., Stéphanie R., Frédérique M., Romuald C., Nathalie P.-L. Merci à tou-te-s celles et ceux vers qui j'ai pu me tourner quand j'avais une question.

Merci à mes premiers co-encadrants en recherche scientifique, Jonathan Webb, puis Maud Rimbault et Yannick Outreman, aux côtés de Julie et Jean-Christophe.

Je tiens également à remercier celles et ceux qui, sans forcément contribuer directement à mon travail, ont participé à faire de ces 3 et quelques années une très belle expérience ! Merci tout particulièrement à Leyli, Maïwenn, Marianne, Théo L., Théo V., Ginette, Yannis, Rémi R., Ségolène, Delphine, Léna et al., ainsi que mes illustres prédécesseur·e-s, les docteurs Mélanie, Ambre et Rémi : merci d'avoir participé à rendre cette expérience de thèse inoubliable et jamais monotone !

Merci à tou-te-s mes ami-e-s de Rennes et de (beaucoup) plus loin, vous êtes une source de joie et de motivation. Merci à mes colocs préféré-e-s pour tout ce qu'on a partagé. Un grand merci à ma famille bretonne, angevine et corse, toujours optimiste et encourageante.

Je suis extrêmement reconnaissante envers mes parents, qui me soutiennent dans tout ce que j'entreprends.

Ma petite sœur Cécile, merci pour tout !

Introduction

Given that almost all eukaryotic organisms reproduce sexually, at least occasionally, sexual reproduction is considered to be an ancestral characteristic of eukaryotes that arose once in their last common ancestor (Bell, 1982). Sexual reproduction is defined as any process that combines genes from more than one source in an organism (Margulis & Sagan, 1990) involving meiosis - the division of a germ cell that gives rise to gametes-, fertilization, and resulting in the production of offspring that are genetically different from the parent(s) (Lively, 2010; Thornhill & Alcock, 2013). Yet there is a great diversity of reproductive modes among eukaryotes, which has long puzzled evolutionary biologists (e.g. Picard et al., 2021). In particular, some eukaryotes have lost sexual reproduction and reproduce asexually, i.e., without sex and resulting in the production of offspring arising from a single individual without fertilization. Automictic parthenogenesis is a form of asexual reproduction that involves meiosis and recombination while organisms using apomictic parthenogenesis do not undergo recombination and produce genetically identical offspring. Automictic or apomictic parthenogenesis has evolved in various ancestrally sexual taxa, such as in plants, rotifers, nematodes, arthropods or reptiles (Bell, 1982). Parthenogenesis is assumed to be rare in multicellular eukaryotes, found in approximately 0.1% of species (Bell, 1982; Simon et al., 2003; White, 1977). However, this estimate is based on the very low proportion of asexuality in vertebrates (Avisé et al., 1992; White, 1977), and it is likely that the occurrence of asexuality in eukaryotes is vastly underestimated. A recent study focused on parthenogenetic arthropods with haplodiploid sex determination (in which males develop from unfertilized eggs and females from fertilized eggs) and showed that the frequency of parthenogenetic lineages ranges from 0–1.5% between orders, but in species-rich genera, parthenogenesis occurs in up to 38% of the species (van der Kooi et al., 2017). Although they are particularly widespread throughout the tree of life, parthenogenetic taxa tend to occupy terminal nodes of phylogenetic trees, suggesting that they are evolutionary dead-ends. Exceptions to this pattern are the *Bdelloidea* rotifers and *Darwinulidae* ostracods, which form a higher taxonomic group with multiple putatively ancient parthenogenetic species (Judson & Normark, 1996; Mark Welch & Meselson, 2000), although there is growing evidence of signature of recombination in these organisms (Laine et al., 2022; Simion et al., 2021; Vakhrusheva et al., 2020).

These independent and frequent sex loss events across various taxa of eukaryotes raise questions about the mechanisms underlying sex loss and the ease with which it can occur, as well as the consequences of such transitions towards asexuality. My thesis focused on these two questions, which I addressed using the aphid model. In this introduction, I will first review what is known about the genetic basis of sex loss in animals, and present the conceptual framework and empirical data regarding the evolutionary consequences of sex loss. I will then introduce the aphid system and its variation in reproductive mode, which makes it particularly well suited to studying these evolutionary biology issues. Finally, I will describe in more detail the purposes of my thesis and how I have attempted to address them.

1. Genetic basis of sex loss in eukaryotes

Transition to asexuality may result from different mechanisms that disrupt meiosis, such as novel mutations, polyploidization, hybridization of closely related lineages (Neiman et al., 2014) but also infection by microorganisms (Werren et al., 2008). Spontaneous transition to obligate asexuality through individual mutations is more likely to occur in cyclically or facultatively asexual species, i.e. species with a life cycle comprising an alternation of sexual and asexual generation (cyclical) or a production of sexual and asexual offspring by the same female (facultative). Then repeated gene flow from essentially obligate asexual lineages into cyclically asexual ones can also make them obligate asexual, a process also referred to as contagious asexuality (Delmotte et al., 2001; Innes & Hebert, 1988). Little is known about the genetic mechanisms underlying the shifts to obligate asexuality. Indeed, obligate parthenogens prevent researchers to characterize the genetic control of asexuality in these species, because classical crosses cannot be established. Yet, there are species that display lineages showing a partial loss of sexual reproduction, which coexist with sexual lineages, thus allowing an identification of the genetic basis of the transition to asexuality using recombination-based approaches. Some of the studies that have investigated this in animals are described in **Table 1** and have shown that the genetic architecture for transitions to obligate asexuality may be simple, involving only a small number of loci. Few studies have gone so far as to identify the location, nature and function of the candidate loci for sex loss. This gap of knowledge prevents testing whether sex loss involves genes with similar functions across taxa and hinders the understanding on how the evolution towards obligate asexuality is constrained. Moreover, the likely divergence in the mechanisms of sex loss among the distant species individually studied makes it difficult to study evolutionary convergence in genes or mechanisms. However, there are taxa that show a partial loss of sex at the inter- or even at the intra-specific level (some species/populations are obligate asexual, others are capable of sexual reproduction), such as stick insects, daphnia or aphids, which may be particularly relevant in this context. In addition to the multiple mechanisms underlying transitions to asexuality, there are different cellular mechanisms by which an unfertilized egg can develop into an offspring, such as apomixis and automixis with many different subforms (Hörandl et al., 2020; Schön et al., 2009). The underlying mechanisms for both the loss of sex and the cytology of asexuality can have profound and diverse consequences for genome evolution (Engelstädter, 2017; Jaron et al., 2021; Parker et al., 2019a), but remain poorly understood in animals.

Table 1. Case studies that have investigating the genetic bases for transitions from sexual or partially sexual to asexual reproduction in arthropods.

Species	Common taxon name	Type of transition	Genetic architecture of obligate asexuality	Asexuality is recessive	Approaches used	Citation
<i>Sitobion avenae</i>	Aphid	Cyclical to obligate asexuality (Apomictic)	2 autosomal loci	Yes	Breeding crosses	Dedryver et al., 2013
<i>Acyrtosiphon pisum</i> (alfalfa race)	Aphid	Cyclical to obligate asexuality (Apomictic)	1 locus on the X chromosome	Yes	Breeding crosses Low-resolution genome scan	Jaquiéry et al., 2014
<i>Brachionus calyciflorus</i>	Monogont rotifer	Cyclical to obligate asexuality (Apomictic)	1 locus	Yes	Breeding crosses	Stelzer et al., 2010
<i>Lysiphlebus fabarum</i>	Wasp	Haplodiploidy to asexuality (Automixis)	1 locus	Yes	Breeding crosses	Sandrock & Vorburger, 2011
<i>Daphnia pulex</i>	Daphnia	Cyclical to obligate asexuality (Apomictic)	1 locus	No	Breeding crosses	Innes & Hebert, 1988
			4 unlinked loci (different chromosomes V, VIII, IX, X)	No	Breeding crosses Genome-wide association studies	Lynch et al., 2008
			An insertion upstream of a copy of a gene (chromosome VIII) encoding a meiotic cohesin		Phylogenetic and whole-genome sequence analysis	Eads et al., 2012
			Polygenic		Breeding crosses Genome-wide association studies	Xu et al., 2015
			3 genes with meiosis-related functions identified among those with allele-specific expression		Comparative transcriptomics	Ye et al., 2021
<i>Apis mellifera capensis</i> (The genetic basis underlying asexuality in this species is disputed)	Honeybee	Haplodiploidy to asexuality (Automixis)	1 locus	Yes	Breeding crosses	Lattorff et al., 2005
			1 candidate locus (<i>th</i> , chromosome 13) coding for a transcription factor	Yes	Breeding crosses Low-resolution genome scan	Lattorff et al., 2007
			<i>th</i> is not associated with asexuality two new polymorphisms found		Breeding crosses	Chapman et al., 2015
			1 locus (<i>th</i>)	Yes	Breeding crosses	Aumer et al., 2017
			1 single nucleotide polymorphism on the identified <i>th</i> locus	No	Genome scan	Aumer et al., 2019
1 gene (chromosome 11) affecting chromosome segregation	Yes	Breeding crosses	Yagound et al., 2020			
<i>Drosophila mercatorum</i>	Fruit fly	Sexual reproduction to facultative parthenogenesis (Automixis)	3 genes (different chromosomes, X, 3rd) involve in parthenogenetic development		Comparative transcriptomics Gene expression manipulation	Sperling et al., 2023
<i>Artemia parthenogenetica</i>	Brine shrimp	Facultative to obligate asexuality (Automixis)	1 locus on the Z chromosome	No	Breeding crosses Genome scan	Elkrewi et al., 2022
			1 gene (<i>Dmc1</i>)	No	Breeding crosses Comparative transcriptomics Knocking-down by RNAi	Boyer et al., 2021 Xu et al., 2023

2. Evolutionary consequences of sex loss in eukaryotes

The loss of sexual reproduction has two major effects: a meiotic recombination arrest (for apomictic parthenogenetic species) and the absence of sexual morphs, for which consequences are predicted on the scale of genomes and gene expression.

2.1. Effect of the lack of genomic recombination in clonal asexual species

2.1.1. Genome evolution in non-recombining asexual species

One theorized advantage of sex and especially of meiotic recombination is that it aids in the clearance of deleterious mutations. In asexual populations, high genetic linkage would ultimately reduce the overall effectiveness of selection. Assuming that most mutations are deleterious and that reverse mutations are rare, this would result in the irreversible accumulation of deleterious mutations in asexual species ("Muller's ratchet"; Muller, 1964). Indeed, due to the lack of recombination, offspring bear at least the same mutational load as their parents. Therefore, both the ratio of nonsynonymous (amino acid changing) to synonymous polymorphism and the ratio of nonsynonymous to synonymous substitutions (also termed dN/dS) are predicted to be higher in asexually than in sexually reproducing organisms. This has only met equivocal empirical validation: an effect of asexuality on the nonsynonymous to synonymous substitution ratio was observed in some animal taxa, as snails (Johnson & Howard, 2007; Neiman et al., 2010) and insects (Bast et al., 2018; Henry et al., 2012; Hollister et al., 2015), but there were no significant differences between this ratio for bdelloid and monogonont rotifers (Mark Welch & Meselson, 2001) or asexual aphid species (Ollivier et al., 2012) and ancient asexual oribatid mites actually showed increased effectiveness of purifying selection, i.e. lower dN/dS (Brandt et al., 2017).

Second, high levels of heterozygosity are expected in asexual diploid species since the two alleles within a locus will theoretically accumulate different mutations over time and diverge in parallel, unless the same mutation occurs in the homologous chromosome or homologous sequences become identical through gene conversion. This phenomenon of allelic divergence is known as the Meselson effect (Mark Welch & Meselson, 2000), yet it has received only limited support in animals. Jaron et al. (2021) found that only parthenogens of hybrid origin were characterized by high heterozygosity levels. Divergence between homologous chromosomes in several animals such as aphids, nematodes or worms has been shown to result from hybridization (Ament-Velásquez et al., 2016; Delmotte et al., 2003; Jaron et al., 2021; Lunt, 2008) or from ancient genome duplication in bdelloid rotifers (Flot et al., 2013; Mark Welch et al., 2008; Nowell et al., 2018). However, the expected haplotype divergence has been observed in studies of asexual *Timema* stick insects and fissiparous *Dugesia* flatworms (Leria et al., 2019; Schwander et al., 2011) and more recently, evidence for the Meselson effect was provided by Brandt et al. (2021) in an asexual oribatid mite species that showed signature of haplotype divergence not caused by hybridization or polyploidization.

Third, given that outcrossing and recombination allow the spread of transposable elements (i.e. DNA sequences, simplified as TEs, that can replicate and change positions within a genome via various mechanisms, Burt & Trivers, 2008) throughout the population, TEs should thus be less frequent in asexual than in sexual populations (Wright & Finnegan, 2001). However, asexual species derive from sexual ancestors, from which they inherit their load of TEs. TEs could therefore continue to impose

costs on their hosts despite being unable to spread to other lineages. Following the abandonment of sex, large asexual populations are expected to purge their load of TEs, thanks to mutation, including deletion, and selection against clones in which active TEs have increased the deleterious load. On the other hand, computer simulations predict that a small population is driven to extinction by a Muller's ratchet-like process of TEs accumulation in the genome (Dolgin & Charlesworth, 2006). Arkhipova and Meselson (2000) compared the presence of transposons within a large body of animal phyla and found that bdelloid rotifers, that are believed to reproduce only asexually for many million years (Mark Welch & Meselson, 2000) – but see above for this debated issue, lack TEs relative to all other tested species. TE numbers were also found to be higher in sexual populations than in asexual populations of *Daphnia pulex* (Schaack et al., 2010). Generally, the abundance of TEs in parthenogenetic animal genomes appears to be lower than in sexual species, but whether this is linked to sex loss remains an open question as TE loads are known to be highly lineage-specific (see Jaron et al., 2021). As pointed out by Glémin and Galtier (2012), the general pattern of transposable elements abundance in asexuals is still unclear.

Other features of asexual genomes that were suggested to be caused by their mode of reproduction (e.g. horizontal gene transfer, palindromes, genomic rearrangements, gene losses, gene family expansions and gene conversion) are mostly lineage-specific and not generally linked to asexuality as indicated by Jaron et al. (2021) who tested these genome features on 26 published animal genomes.

2.1.2. Gene expression evolution in non-recombining asexual species

The lack of recombination in asexual genomes could also affect gene expression evolution. First, the patterns of gene expression in sexual and asexual lineages should reflect results found on the genetic sequences and be an indication of the overall less efficient selection in asexual lineages. It is still unclear how the selective pressures faced by asexual lineages affect the evolution of gene expression and observed traits. It could be expected that, if selection is less efficient on asexual than on sexual lineages, gene expression in asexual lineages would progressively drift away from its optimum in a random manner and we would thus potentially observe a higher variance in gene expression between asexual than between sexual lineages. Second, in diploid species, there can be a transcriptional imbalance between certain alleles, known as allele-specific expression. Gene expression can be cis-regulated by enhancers that regulate gene copy located next to them (on the same chromosome) and not the copy on the homologous chromosome. Recent population genetic theory showed that the evolution of regulatory regions can be strongly influenced by reproductive systems (Fyon & Lenormand, 2018). In clonal asexual species displaying high levels of heterozygosity (most likely apomictic species that do not recombine), if homologous chromosomes are sufficiently isolated from one another, homologous *cis*-regulatory regions are expected to diverge, which would lead to haploidization of expression (Fyon & Lenormand, 2018). The hypothesis of expression divergence between homologous chromosomes in clonal animals is difficult to test. Ye et al. (2021) found that obligate asexual clones have significantly more genes experiencing allele-specific expression than cyclically asexual clones in *Daphnia pulex*, but OP clones derived from hybridization between CP *D. pulex* and its sister species *Daphnia pulicaria* so unbalanced expression of alleles between OP and CP clones is more likely to be due to hybridization rather than lack of recombination.

2.2. Effect of the absence of sexual morphs

The shift from a sexually reproducing species with males and females to an asexual all-female species has major evolutionary consequences. First, the absence of sexual morphs relaxes selection on the morphs that are no longer produced. Second, it also reduces genetic conflict that existed between males and females due to the constraint of a shared genome. Indeed, in sexually-reproducing species, two morphs, males and females, share the same genome (except for sex chromosomes). If a trait is controlled by the same genes in males and females, neither sex can reach its optimal fitness (Parker, 1979). This leads to intra-locus sexual conflict, the resolution of which may involve differential gene expression according to sex. Yet unsolved or partially unsolved conflicts between sexes may remain, leading to suboptimal patterns of gene expression in each sex (**Figure 1**).

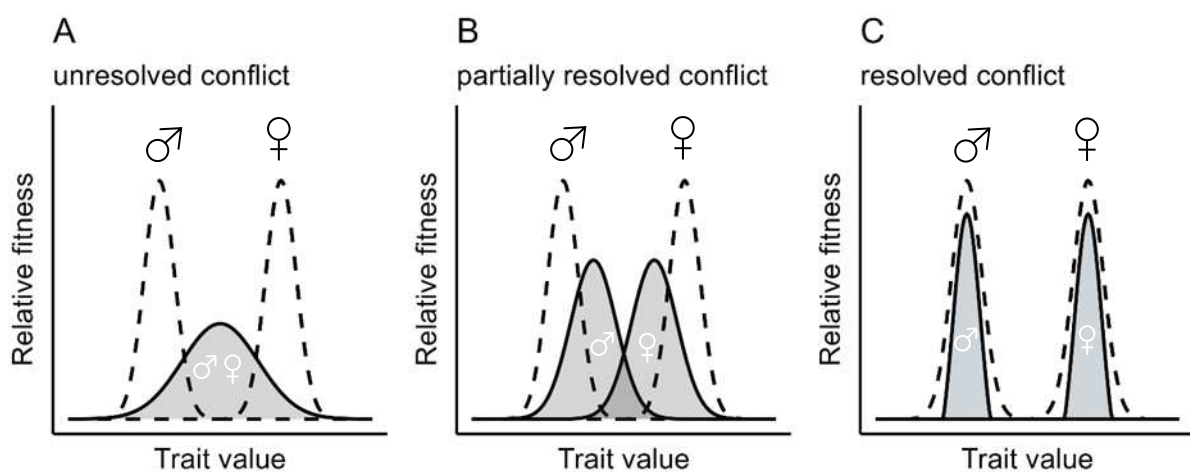


Figure 1. Relationships between sexual dimorphism and intra-locus sexual conflict. Shaded areas indicate phenotypic distributions for a hypothetical quantitative trait, and dashed lines indicate fitness functions for that trait. **(A)** Intra-locus conflict occurs when fitness optima for a trait, but not trait expression, differ between males and females. **(B)** Sexual dimorphism can occur without full resolution of intra-locus conflict, which would occur if **(C)** the trait expression distributions fell exactly under the fitness optima distributions. Adapted from Cox and Calsbeek (2009) and Gamble and Calsbeek (2021).

Sexual antagonism is arguably one major force leading to the currently observed balance between the sex-biased expression of genes carried by both males and females (Veltso et al., 2017). Indeed, in genetic sex-determination systems, male and female genomes differ only by sex chromosomes but the level of expression of many genes, including genes on the autosomes (around 75% of the genes in *Drosophila melanogaster* genome, Assis et al., 2012) depends on the sex of the individuals. Genes that are subject to different selection pressures depending on the sex may evolve sex-specific expression and are called sex-biased genes (Ellegren & Parsch, 2007). Depending on the sex that shows the highest expression of these genes, they are either female-biased or male-biased genes.

Here I will consider the consequences of two aspects of the absence of sexual morphs, relaxed selection on the absent sexual morphs and reduced intra-locus conflict, on genomic sequence and gene expression evolution.

2.2.1. Gene sequence evolution in asexuals

Genes that were previously required for sexual reproduction are expected to decay in asexual lineages. If they are not expressed anymore, as it could be the case for male-limited genes (only expressed in males) in asexual all-female species, they may slowly decay as a consequence of neutral mutation accumulation. In contrast, some genes encoding sexual reproduction functions may undergo strong selection for loss if their activity becomes an unnecessary metabolic burden for asexual morphs. This would result in a high nonsynonymous to synonymous substitution and polymorphism rates (dN/dS and pN/pS ratio). The molecular evolution of sex-biased genes after sex loss has been the subject of very few studies. In an asexual parasitoid wasp, in which male-specific sexual traits are not expressed and can only decay through neutral mutations, Kraaijeveld et al. (2016) showed that genes underlying these traits do indeed accumulate deleterious mutations. In the pea aphid, a species that produces multiple morphs at various frequency during its life cycle, genes with expression biased towards rarer morphs (sexual females and males) show faster rates of evolution than genes expressed in the more common morph (asexual females), most likely because of relaxed purifying selection (Purandare et al., 2014). Therefore, if the amount of time a gene spends being expressed in a morph is associated with its rate of evolution, we can speculate that genes expressed in sexual morphs that are completely absent in parthenogenetic lineages would show fast rate of evolution.

2.2.2. Gene expression evolution in asexuals

Shifts to obligate asexuality render obsolete the sexual selective pressures on genomes, especially on sexually antagonistic genes, and therefore intra-locus sexual conflicts. Asexuality is therefore expected to overcome this constrained evolution of the expression of the previously sexually antagonistic genes. Any asexual lineage that invests more in female functions compared to sexual species - at the expense of males that are absent and therefore do not constrain gene expression in females - should be favored by selection. Thus, we would expect to see higher expression of female-biased genes and lower expression of male-biased genes in asexual compared to sexual species. Few studies have addressed the evolution of sex-biased gene expression after a relaxation of intra-locus conflict following sex loss and theoretical predictions tend not to be fulfilled. Instead of a feminization, Parker et al. (2019b) found evidence for masculinization of sex-biased gene expression in asexual females when comparing expression of pairs of asexual and closely related sexual species of *Timema* stick insects. On the other hand, no shift in expression in sex-biased genes after sex loss was observed in *Artemia* brine shrimp asexual species compared to sexual species (Huylmans et al., 2021). It hence seems that there is no general pattern as to how the transcriptome evolves after the cessation of sexual conflict following sex loss. Findings are sometimes difficult to interpret in such studies, as the shift from obligate sexual to obligate asexual reproduction implies the evolution of a new asexual morph whose optimal gene expression is likely to differ from that of the sexual females with which the transcriptomes are compared. One promising approach would be to compare the expression of genes previously involved in intra-locus conflict between asexual and sexual populations that have already evolved parthenogenesis, as is the case in cyclically or facultatively asexual species.

3. Aphids as models for evolutionary studies of asexuality

Aphids, which have evolved partial asexuality long ago and have then recently lost multiple times the sexual phase, are of particular interest for studying the evolution of asexuality. The common sexual ancestor of aphids acquired the ability to alternate between parthenogenetic and sexual reproduction ~250 million years ago (Dixon, 1998; Hales et al., 1997). This alternation of many asexual (*via* apomictic parthenogenesis) and one sexual generation a year is called cyclical parthenogenesis (CP), and typically involves three reproductive morphs: viviparous parthenogenetic females, which are present for the main part of the aphid life cycle, males and oviparous sexual females, both of which are produced once a year (**Figure 2**). Despite the relatively low frequency of sexual generations in CP aphids (once every 10 to 15 generations), it is enough to result in genetic parameters (diversity and heterozygosity) equivalent to that of an organism that would reproduce only sexually (Delmotte et al., 2002; Rouger et al., 2016; Simon et al., 2002).

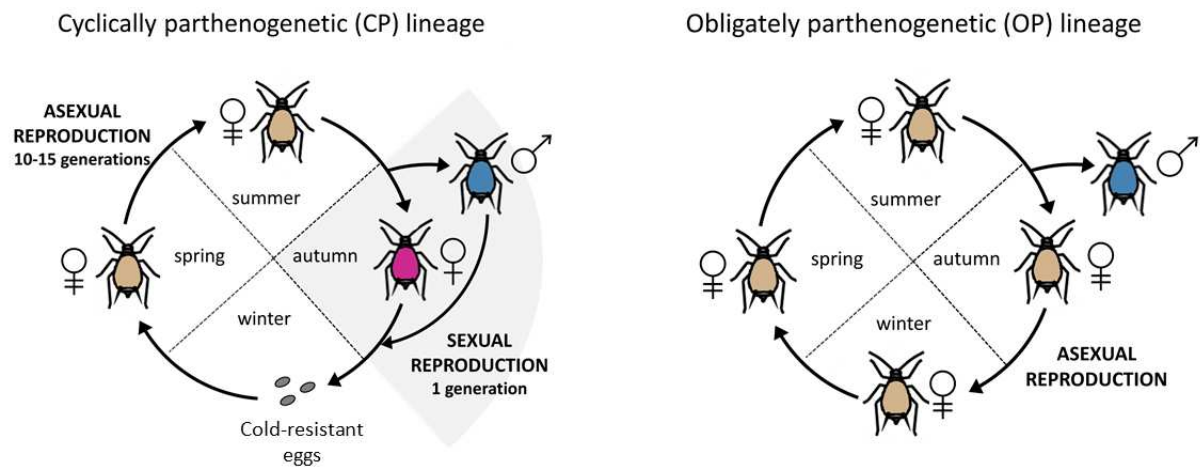


Figure 2. Annual life cycle of lineages of the pea aphid *Acyrthosiphon pisum* that differ in their investment in sexual reproduction. In CP lineages, overwintering eggs produced by sexual reproduction are cold-resistant and diapause during winter, before giving birth to asexual females in the following spring. After several cycles of parthenogenesis, the shortening of the photoperiod in autumn induces the production of sexual females and males. In OP lineages, aphids reproduce by viviparous parthenogenesis throughout the year, with unequal production of males but no sexual females. As parthenogenetic females are less resistant to cold than the eggs produced by CP lineages, CP lineages are favored by selection in cold-winter regions compared to OP lineages. However, selection favors continued parthenogenetic reproduction in mild-winter regions, which gives OP lineages a demographic advantage over CP lineages. This results in a geographic cline of both types of lineages.

Remarkably, some aphid lineages have secondarily lost the ability to reproduce sexually, becoming obligate parthenogens (OP, **Figure 2**). OP lineages do not produce sexual females but they produce parthenogenetic females and some males. Since the egg resulting from sexual reproduction is the sole cold-resistant stage in aphids, only CP lineages can survive in cold-winter regions (Simon et al., 2002, 2010). Contrastingly, OP lineages cannot produce eggs (due to the absence of sexual females), they are thus frost-sensitive and are found in mild-winter regions where their ability to maintain parthenogenetic development in winter gives them a demographic advantage over CP

lineages (Simon et al., 2002). This geographic separation of CP and OP lineages due to climate has been shown for several aphid species (Blackman, 1974; Dedryver et al., 2001; Papura et al., 2003; Simon et al., 1999; Vorburger et al., 2003). As a result, OP males have low opportunity to find mating partners (i.e., sexual females from CP lineages) and to pass on their genes to the next generation.

4. Thesis objectives

First, we wanted to decipher the genetic architecture of the loss of sex and investigate its evolutionary convergence in aphids to shed light on why obligate asexuality is so widespread in this group. Previous to my PhD, the genetic basis of the shift to obligate parthenogenesis has been studied with quantitative genetic and population genomic approaches by Jaquiéry et al. (2014) in the alfalfa race of the pea aphid complex *Acyrtosiphon pisum*. Obligate parthenogenesis, a recessive trait, was shown to be controlled by a large genomic region of 10 cM on the X chromosome. The study of other aphid populations will provide a better understanding of the genetic mechanisms underlying sex loss and of its evolutionary history in a group characterized by numerous independent losses of sexual reproduction. In the **first chapter**, we aimed at characterizing more precisely this region and the underlying genes, with higher resolution genome scan based on pool-seq data of OP and CP populations of the alfalfa race of *A. pisum*. We then set out to identify the genetic basis of sex loss along a continuum of divergence from two other closely related races of the pea aphid to a more divergent species, the green peach aphid, *Myzus persicae* (**chapter 2**). The pea aphid complex is indeed a very relevant system to study the convergence of the genetic basis of sex loss at the intra-specific level, as it comprises at least 15 races adapted to different legume hosts that have recently diverged (radiation was estimated to be around 18,000-47,000 years old, Peccoud et al., 2009) and for which variation in reproductive mode is observed. Among these races, we worked with the pea and clover races of *A. pisum* for which pooled genomic data were available and consisted of several pools of individuals from genetically unique lineages collected in France and segregated according to their reproductive mode. More practically, we used genome-wide association approach to detect associations between allele frequency and reproductive mode. To understand whether the identified genomic regions are common or species-specific, we then used genomic data of another species *M. persicae* that diverged from *A. pisum* 22 million years ago (Mathers et al., 2021) and for which reproductive mode variation was suggested to be controlled by a limited set of loci according to results of breeding crosses performed by Blackman (1972). Sympatric lineages sequenced in pools of different reproductive mode (data reanalyzed from Roy et al., 2013, 2022), as well as published individually sequenced clones with global origin (Singh et al., 2021) were used for this study and OP and CP populations were compared to detect any highly differentiated genomic regions between reproductive modes. Using genome scan approaches, we identified candidate regions associated with the reproductive mode in this species and then searched for similarities in sequence, localization or gene content between the candidate regions of *A. pisum* and *M. persicae*.

The second part of my PhD focused on the phenotypic and transcriptomic consequences of sex loss. Sex loss can result in genes with functions related to sexual reproduction to lose their adaptive value under asexuality. Here, we wanted to look for evidence of changes in traits and gene expression patterns resulting from shifts in evolutionary forces associated with sex loss. Investigating the evolutionary consequences of sex loss can be complicated: for example, interspecific comparisons

increase the likelihood that observations are due to divergence between species rather than changes in evolutionary forces associated with sex loss, and the study of single asexual lineages may only reveal lineage-specific results. In this context, the pea aphid is a convenient system as it shows an intraspecific variation in reproductive mode associated with climate, with cold-resistant CP lineages displaying the ancestral reproductive mode and cold-sensitive OP lineages producing no sexual females but still males. As OP and CP lineages are geographically separated, OP males have limited reproductive opportunity and are hence expected to be subject to relaxed selection. This relaxation of selection on males may lead to a lower reproductive success, as it is expected that genes predominantly or exclusively expressed in males (i.e., male-biased genes) accumulate more deleterious mutations in OP compared to CP lineages. In addition, unlike CP genomes which recombine and mix, each OP genome evolves independently and is expected to accumulate broadly deleterious mutations with the age of the last sexual event (Muller, 1964). This could also lead to greater variance in male reproductive success between OP lineages compared to CP lineages. In my thesis, I studied the consequences of relaxed selection on sexual traits in asexually reproducing lineages in the clover race of *A. pisum*. In **chapter 3**, I compared the reproductive success of males produced by 12 CP lineages and 12 OP lineages, by carrying out mating experiments with oviparous sexual females in which these males were either alone or in the presence of (and thus in competition with) other males. To expand this analysis to other functional traits, we considered the evolutionary consequences of sex loss at the transcriptomic scale. Sexual conflict no longer exists in asexual lineages, making them relevant systems for understanding the extent to which sexual conflict drives gene expression. Aphids are particularly well-suited organisms for addressing this question because cyclically parthenogenetic lineages already can produce asexual along with sexual morphs, which simplifies expectations about the evolution of gene expression. First, assuming that OP males are under relaxed selection and sexual females are not produced in OP lineages, only the parthenogenetic female is expected to be under strong selection in OP lineages. Conflicts between morphs are hence relaxed and we expect changes in expression of morph-biased genes that may have been involved in an intra-locus conflict that no longer exists in OP lineages. Second, the lack of mixing between OP lineages (due to the absence of recombination) implies that OP lineages will evolve independently from each other and accumulate deleterious mutations. Therefore, a higher variance in gene expression within morphs produced by OP lineages is expected relative to CP lineages. To test these predictions, we performed whole-body transcriptomic sequencing for the morphs in 4 OP and 4 CP *A. pisum* lineages of the clover race (**chapter 4**). Morph-biased genes were identified with independent transcriptomic data and we tested for gene expression differences between CP and OP morphs.

5. References

- Ament-Velázquez, S. L., Figuet, E., Ballenghien, M., Zattara, E. E., Norenburg, J. L., Fernández-Álvarez, F. A., Bierne, J., Bierne, N., & Galtier, N. (2016). Population genomics of sexual and asexual lineages in fissiparous ribbon worms (Lineus, *Nemertea*): Hybridization, polyploidy and the Meselson effect. *Molecular Ecology*, *25*(14), 3356–3369. <https://doi.org/10.1111/mec.13717>
- Arkipova, I., & Meselson, M. (2000). Transposable elements in sexual and ancient asexual taxa. *Proceedings of the National Academy of Sciences*, *97*(26), 14473–14477. <https://doi.org/10.1073/pnas.97.26.14473>
- Assis, R., Zhou, Q., & Bachtrog, D. (2012). Sex-biased transcriptome evolution in *Drosophila*. *Genome Biology and Evolution*, *4*(11), 1189–1200. <https://doi.org/10.1093/gbe/evs093>
- Aumer, D., Allsopp, M. H., Lattorff, H. M. G., Moritz, R. F. A., & Jarosch-Perlow, A. (2017). Thelytoky in Cape honeybees (*Apis mellifera capensis*) is controlled by a single recessive locus. *Apidologie*, *48*(3), 401–410. <https://doi.org/10.1007/s13592-016-0484-0>
- Aumer, D., Stolle, E., Allsopp, M., Mumoki, F., Pirk, C. W. W., & Moritz, R. F. A. (2019). A single SNP turns a social honey bee (*Apis mellifera*) worker into a selfish parasite. *Molecular Biology and Evolution*, *36*(3), 516–526. <https://doi.org/10.1093/molbev/msy232>
- Avise, J. C., Quattro, J. M., & Vrijenhoek, R. C. (1992). Molecular clones within organismal clones. In M. K. Hecht, B. Wallace, & R. J. Macintyre (Eds.), *Evolutionary Biology: Volume 26* (pp. 225–246). Springer US. https://doi.org/10.1007/978-1-4615-3336-8_6
- Bast, J., Parker, D. J., Dumas, Z., Jalvingh, K. M., Tran Van, P., Jaron, K. S., Figuet, E., Brandt, A., Galtier, N., & Schwander, T. (2018). Consequences of asexuality in natural populations: insights from stick insects. *Molecular Biology and Evolution*, *35*(7), 1668–1677. <https://doi.org/10.1093/molbev/msy058>
- Bell, G. (1982). *The masterpiece of nature: The evolution and genetics of sexuality*. University of California Press.
- Blackman, R. L. (1972). The inheritance of life-cycle differences in *Myzus persicae* (Sulz.) (*Hem., Aphididae*). *Cambridge University Press*, *62*(2), 281–294. <https://doi.org/10.1017/S0007485300047726>
- Blackman, R. L. (1974). Life-cycle variation of *Myzus persicae* (Sulz.) (*Hom., Aphididae*) in different parts of the world, in relation to genotype and environment. *Bulletin of Entomological Research*, *63*(4), 595–607. <https://doi.org/10.1017/S0007485300047830>
- Boyer, L., Jabbour-Zahab, R., Mosna, M., Haag, C. R., & Lenormand, T. (2021). Not so clonal asexuals: Unraveling the secret sex life of *Artemia parthenogenetica*. *Evolution Letters*, *5*(2), 164–174. <https://doi.org/10.1002/evl3.216>
- Brandt, A., Schaefer, I., Glanz, J., Schwander, T., Maraun, M., Scheu, S., & Bast, J. (2017). Effective purifying selection in ancient asexual oribatid mites. *Nature Communications*, *8*(1), 1. <https://doi.org/10.1038/s41467-017-01002-8>
- Burt, A., & Trivers, R. (2008). Genes in conflict: the biology of selfish genetic elements. In *Genes in Conflict*. Harvard University Press. <https://doi.org/10.4159/9780674029118>
- Chapman, N. C., Beekman, M., Allsopp, M. H., Rinderer, T. E., Lim, J., Oxley, P. R., & Oldroyd, B. P. (2015). Inheritance of thelytoky in the honey bee *Apis mellifera capensis*. *Heredity*, *114*(6), 6. <https://doi.org/10.1038/hdy.2014.127>
- Cox, R. M., & Calsbeek, R. (2009). Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *The American Naturalist*, *173*(2), 176–187. <https://doi.org/10.1086/595841>
- Dedryver, C.-A., Hullé, M., Le Gallic, J.-F., Caillaud, M. C., & Simon, J.-C. (2001). Coexistence in space and time of sexual and asexual populations of the cereal aphid *Sitobion avenae*. *Oecologia*, *128*(3), 379–388. <https://doi.org/10.1007/s004420100674>
- Dedryver, C.-A., Le Gallic, J.-F., Mahéo, F., Simon, J.-C., & Dedryver, F. (2013). The genetics of obligate parthenogenesis in an aphid species and its consequences for the maintenance of alternative reproductive modes. *Heredity*, *110*(1), 1. <https://doi.org/10.1038/hdy.2012.57>
- Delmotte, F., Leterme, N., Bonhomme, J., Rispe, C., & Simon, J.-C. (2001). Multiple routes to asexuality in an aphid species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1483), 2291–2299. <https://doi.org/10.1098/rspb.2001.1778>

- Delmotte, F., Leterme, N., Gauthier, J.-P., Rispe, C., & Simon, J.-C. (2002). Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology*, *11*(4), 711–723. <https://doi.org/10.1046/j.1365-294X.2002.01478.x>
- Delmotte, F., Sabater-Muñoz, B., Prunier-Leterme, N., Latorre, A., Sunnucks, P., Rispe, C., & Simon, J.-C. (2003). Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. *Evolution*, *57*(6), 1291–1303. <https://doi.org/10.1111/j.0014-3820.2003.tb00337.x>
- Dixon, A. F. G. (1998). *Aphid Ecology* (2nd ed.). Chapman & Hall.
- Dolgin, E. S., & Charlesworth, B. (2006). The fate of transposable elements in asexual populations. *Genetics*, *174*(2), 817–827. <https://doi.org/10.1534/genetics.106.060434>
- Eads, B. D., Tsuchiya, D., Andrews, J., Lynch, M., & Zolan, M. E. (2012). The spread of a transposon insertion in *Rec8* is associated with obligate asexuality in *Daphnia*. *Proceedings of the National Academy of Sciences*, *109*(3), 858–863. <https://doi.org/10.1073/pnas.1119667109>
- Elkrewi, M., Khauratovich, U., Toups, M. A., Bett, V. K., Mrnjavac, A., Macon, A., Fraise, C., Sax, L., Huylmans, A. K., Hontoria, F., & Vicoso, B. (2022). ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp. *Genetics*, iyac123. <https://doi.org/10.1093/genetics/iyac123>
- Ellegren, H., & Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, *8*(9), 9. <https://doi.org/10.1038/nrg2167>
- Engelstädter, J. (2017). Asexual but not clonal: evolutionary processes in automictic populations. *Genetics*, *206*(2), 993–1009. <https://doi.org/10.1534/genetics.116.196873>
- Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E. G. J., Hejnol, A., Henrissat, B., Koszul, R., Aury, J.-M., Barbe, V., Barthélémy, R.-M., Bast, J., Bazykin, G. A., Chabrol, O., Couloux, A., Da Rocha, M., Da Silva, C., Gladyshev, E., ... Van Doninck, K. (2013). Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature*, *500*(7463), 7463. <https://doi.org/10.1038/nature12326>
- Fyon, F., & Lenormand, T. (2018). Cis-regulator runaway and divergence in asexuals. *Evolution*, *72*(3), 426–439. <https://doi.org/10.1111/evo.13424>
- Gamble, M. M., & Calsbeek, R. G. (2021). A potential role for restricted intertactical heritability in preventing intralocus conflict. *Evolutionary Applications*, *14*(11), 2576–2590. <https://doi.org/10.1111/eva.13292>
- Glémin, S., & Galtier, N. (2012). Genome evolution in outcrossing versus selfing versus asexual species. In M. Anisimova (Ed.), *Evolutionary Genomics: Statistical and Computational Methods, Volume 1* (pp. 311–335). Humana Press. https://doi.org/10.1007/978-1-61779-582-4_11
- Hales, D. F., Tomiuk, J., Woehrmann, K., & Sunnucks, P. (1997). Evolutionary and genetic aspects of aphid biology: A review. *EJE*, *94*(1), 1–55.
- Henry, L., Schwander, T., & Crespi, B. J. (2012). Deleterious mutation accumulation in asexual *Timema* Stick Insects. *Molecular Biology and Evolution*, *29*(1), 401–408. <https://doi.org/10.1093/molbev/msr237>
- Hollister, J. D., Greiner, S., Wang, W., Wang, J., Zhang, Y., Wong, G. K.-S., Wright, S. I., & Johnson, M. T. J. (2015). Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. *Molecular Biology and Evolution*, *32*(4), 896–905. <https://doi.org/10.1093/molbev/msu345>
- Hörandl, E., Bast, J., Brandt, A., Scheu, S., Bleidorn, C., Cordellier, M., Nowrousian, M., Begerow, D., Sturm, A., Verhoeven, K., Boenigk, J., Friedl, T., & Dunthorn, M. (2020). *Genome Evolution of Asexual Organisms and the Paradox of Sex in Eukaryotes*.
- Huylmans, A. K., Macon, A., Hontoria, F., & Vicoso, B. (2021). Transitions to asexuality and evolution of gene expression in *Artemia* brine shrimp. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1959), 20211720. <https://doi.org/10.1098/rspb.2021.1720>
- Innes, D. J., & Hebert, P. D. N. (1988). The origin and genetic basis of obligate parthenogenesis in *Daphnia pulex*. *Evolution*, *42*(5), 1024–1035. <https://doi.org/10.1111/j.1558-5646.1988.tb02521.x>
- Jaquiéry, J., Stoeckel, S., Larose, C., Nouhaud, P., Rispe, C., Miezuet, L., Bonhomme, J., Mahéo, F., Legeai, F., Gauthier, J.-P., Prunier-Leterme, N., Tagu, D., & Simon, J.-C. (2014). Genetic control of contagious asexuality in the pea aphid. *PLOS Genetics*, *10*(12), 12. <https://doi.org/10.1371/journal.pgen.1004838>
- Jaron, K. S., Bast, J., Nowell, R. W., Ranallo-Benavidez, T. R., Robinson-Rechavi, M., & Schwander, T. (2021). Genomic features of parthenogenetic animals. *Journal of Heredity*, *112*(1), 19–33. <https://doi.org/10.1093/jhered/esaa031>

- Johnson, S. G., & Howard, R. S. (2007). Contrasting patterns of synonymous and nonsynonymous sequence evolution in asexual and sexual freshwater snail lineages. *Evolution*, *61*(11), 2728–2735. <https://doi.org/10.1111/j.1558-5646.2007.00233.x>
- Judson, O. P., & Normark, B. B. (1996). Ancient asexual scandals. *Trends in Ecology & Evolution*, *2 Special issue: Ecology and evolution of sexual reproduction*(11), 41–46.
- Kraaijeveld, K., Anvar, S. Y., Frank, J., Schmitz, A., Bast, J., Wilbrandt, J., Petersen, M., Ziesmann, T., Niehuis, O., de Knijff, P., den Dunnen, J. T., & Ellers, J. (2016). Decay of sexual trait genes in an asexual parasitoid wasp. *Genome Biology and Evolution*, *8*(12), 12. <https://doi.org/10.1093/gbe/evw273>
- Laine, V. N., Sackton, T. B., & Meselson, M. (2022). Genomic signature of sexual reproduction in the bdelloid rotifer *Macrotrachella quadricornifera*. *Genetics*, *220*(2), iyab221. <https://doi.org/10.1093/genetics/iyab221>
- Lattorff, H. M. G., Moritz, R. F. A., Crewe, R. M., & Solignac, M. (2007). Control of reproductive dominance by the thelytoky gene in honeybees. *Biology Letters*, *3*(3), 292–295. <https://doi.org/10.1098/rsbl.2007.0083>
- Lattorff, H. M. G., Moritz, R. F. A., & Fuchs, S. (2005). A single locus determines thelytokous parthenogenesis of laying honeybee workers (*Apis mellifera capensis*). *Heredity*, *94*(5), 5. <https://doi.org/10.1038/sj.hdy.6800654>
- Leria, L., Vila-Farré, M., Solà, E., & Riutort, M. (2019). Outstanding intraindividual genetic diversity in fissiparous planarians (*Dugesia*, Platyhelminthes) with facultative sex. *BMC Evolutionary Biology*, *19*(1), 130. <https://doi.org/10.1186/s12862-019-1440-1>
- Lively, C. M. (2010). A Review of red queen models for the persistence of obligate sexual reproduction. *Journal of Heredity*, *101*(suppl_1), S13–S20. <https://doi.org/10.1093/jhered/esq010>
- Lunt, D. H. (2008). Genetic tests of ancient asexuality in Root Knot Nematodes reveal recent hybrid origins. *BMC Evolutionary Biology*, *8*(1), 194. <https://doi.org/10.1186/1471-2148-8-194>
- Lynch, M., Seyfert, A., Eads, B., & Williams, E. (2008). Localization of the genetic determinants of meiosis suppression in *Daphnia pulex*. *Genetics*, *180*(1), 317–327. <https://doi.org/10.1534/genetics.107.084657>
- Margulis, L., & Sagan, D. (1990). *Origins of Sex: Three Billion Years of Genetic Recombination*. Yale University Press.
- Mark Welch, D. B., Mark Welch, J. L., & Meselson, M. (2008). Evidence for degenerate tetraploidy in bdelloid rotifers. *Proceedings of the National Academy of Sciences*, *105*(13), 5145–5149. <https://doi.org/10.1073/pnas.0800972105>
- Mark Welch, D. B., & Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, *288*(5469), 1211–1215. <https://doi.org/10.1126/science.288.5469.1211>
- Mark Welch, D. B., & Meselson, M. (2001). Rates of nucleotide substitution in sexual and anciently asexual rotifers. *Proceedings of the National Academy of Sciences*, *98*(12), 6720–6724. <https://doi.org/10.1073/pnas.111144598>
- Mathers, T. C., Wouters, R. H. M., 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>
- Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *1*(1), 2–9. [https://doi.org/10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8)
- Neiman, M., Hehman, G., Miller, J. T., Logsdon, J. M., Jr, & Taylor, D. R. (2010). Accelerated mutation accumulation in asexual lineages of a freshwater snail. *Molecular Biology and Evolution*, *27*(4), 954–963. <https://doi.org/10.1093/molbev/msp300>
- Neiman, M., Sharbel, T. F., & Schwander, T. (2014). Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. *Journal of Evolutionary Biology*, *27*(7), 1346–1359. <https://doi.org/10.1111/jeb.12357>
- Nowell, R. W., Almeida, P., Wilson, C. G., Smith, T. P., Fontaneto, D., Crisp, A., Micklem, G., Tunnacliffe, A., Boschetti, C., & Barraclough, T. G. (2018). Comparative genomics of bdelloid rotifers: Insights from desiccating and nondesiccating species. *PLOS Biology*, *16*(4), e2004830. <https://doi.org/10.1371/journal.pbio.2004830>
- Ollivier, M., Gabaldón, T., Poulain, J., Gavory, F., Leterme, N., Gauthier, J.-P., Legeai, F., Tagu, D., Simon, J. C., & Rispe, C. (2012). Comparison of gene repertoires and patterns of evolutionary rates in eight aphid species that




- differ by reproductive mode. *Genome Biology and Evolution*, 4(2), 155–167. <https://doi.org/10.1093/gbe/evr140>
- Papura, D., Simon, J.-C., Halkett, F., Delmotte, F., Le Gallic, J.-F., & Dedryver, C.-A. (2003). Predominance of sexual reproduction in Romanian populations of the aphid *Sitobion avenae* inferred from phenotypic and genetic structure. *Heredity*, 90, 5. <https://doi.org/10.1038/sj.hdy.6800262>
- Parker, D. J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M., & Schwander, T. (2019a). Repeated evolution of asexuality involves convergent gene expression changes. *Molecular Biology and Evolution*, 36(2), 350–364. <https://doi.org/10.1093/molbev/msy217>
- Parker, D. J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M., & Schwander, T. (2019b). Sex-biased gene expression is repeatedly masculinized in asexual females. *Nature Communications*, 10(1), 1. <https://doi.org/10.1038/s41467-019-12659-8>
- Parker, G. A. (1979). *Sexual Selection and Reproductive Competition in Insects* (Vol. 123, p. 166).
- Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences*, 106(18), 7495–7500. <https://doi.org/10.1073/pnas.0811117106>
- Picard, M. A. L., Vicoso, B., Bertrand, S., & Escriva, H. (2021). Diversity of modes of reproduction and sex determination systems in invertebrates, and the putative contribution of genetic conflict. *Genes*, 12(8), 8. <https://doi.org/10.3390/genes12081136>
- Purandare, S. R., Bickel, R. D., Jaquiere, J., Rispe, C., & Brisson, J. A. (2014). Accelerated evolution of morph-biased genes in pea aphids. *Molecular Biology and Evolution*, 31(8), 2073–2083. <https://doi.org/10.1093/molbev/msu149>
- Rouger, R., Reichel, K., Malrieu, F., Masson, J. P., & Stoeckel, S. (2016). Effects of complex life cycles on genetic diversity: Cyclical parthenogenesis. *Heredity*, 117(5), 5. <https://doi.org/10.1038/hdy.2016.52>
- Roy, L., Barrès, B., Capderrey, C., Mahéo, F., Micoud, A., Hullé, M., & Simon, J.-C. (2022). Host plants and insecticides shape the evolution of genetic and clonal diversity in a major aphid crop pest. *Evolutionary Applications*, 15(10), 1653–1669. <https://doi.org/10.1111/eva.13417>
- Roy, L., Fontaine, S., Caddoux, L., Micoud, A., & Simon, J.-C. (2013). Dramatic changes in the genotypic frequencies of target insecticide resistance in french populations of *Myzus persicae* (Hemiptera: Aphididae) over the last decade. *Journal of Economic Entomology*, 106(4), 1838–1847. <https://doi.org/10.1603/EC12475>
- Sandrock, C., & Vorburger, C. (2011). Single-locus recessive inheritance of asexual reproduction in a parasitoid wasp. *Current Biology*, 21(5), 433–437. <https://doi.org/10.1016/j.cub.2011.01.070>
- Schaack, S., Pritham, E. J., Wolf, A., & Lynch, M. (2010). DNA transposon dynamics in populations of *Daphnia pulex* with and without sex. *Proceedings of the Royal Society B: Biological Sciences*, 277(1692), 2381–2387. <https://doi.org/10.1098/rspb.2009.2253>
- Schön, I., Martens, K., & Dijk, P. (Eds.). (2009). *Lost Sex: The Evolutionary Biology of Parthenogenesis*. Springer Netherlands. <https://doi.org/10.1007/978-90-481-2770-2>
- Schwander, T., Henry, L., & Crespi, B. J. (2011). Molecular evidence for ancient asexuality in *Timema* stick insects. *Current Biology*, 21(13), 1129–1134. <https://doi.org/10.1016/j.cub.2011.05.026>
- Simion, P., Narayan, J., Houtain, A., Derzelle, A., Baudry, L., Nicolas, E., Arora, R., Cariou, M., Cruaud, C., Gaudray, F. R., Gilbert, C., Guiguelmoni, N., Hespels, B., Kozłowski, D. K. L., Labadie, K., Limasset, A., Lliros, M., Marbouty, M., Terwagne, M., ... Van Doninck, K. (2021). Chromosome-level genome assembly reveals homologous chromosomes and recombination in asexual rotifer *Adineta vaga*. *Science Advances*, 7(41), eabg4216. <https://doi.org/10.1126/sciadv.abg4216>
- Simon, J.-C., Baumann, S., Sunnucks, P., Hebert, P. D. N., Pierre, J.-S., Le Gallic, J.-F., & Dedryver, C.-A. (1999). Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, 8(4), 531–545. <https://doi.org/10.1046/j.1365-294x.1999.00583.x>
- Simon, J.-C., Delmotte, F., Rispe, C., & Crease, T. (2003). Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society*, 79(1), 151–163. <https://doi.org/10.1046/j.1095-8312.2003.00175.x>
- Simon, J.-C., Rispe, C., & Sunnucks, P. (2002). Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, 17(1), 1. [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), 6. <https://doi.org/10.1016/j.crv.2010.03.003>
- Singh, K. S., Cordeiro, E. M. G., Troczka, B. J., Pym, A., Mackisack, J., Mathers, T. C., Duarte, A., Legeai, F., Robin, S., Bielza, P., Burrack, H. J., Charaabi, K., Denholm, I., Figueroa, C. C., French-Constant, R. H., Jander, G., Margaritopoulos, J. T., Mazzoni, E., Nauen, R., ... 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), 1. <https://doi.org/10.1038/s42003-021-02373-x>
- Sperling, A. L., Fabian, D. K., Garrison, E., & Glover, D. M. (2023). A genetic basis for facultative parthenogenesis in *Drosophila*. *Current Biology*, 33(17), 3545–3560.e13. <https://doi.org/10.1016/j.cub.2023.07.006>
- Stelzer, C.-P., Schmidt, J., Wiedlroither, A., & Riss, S. (2010). Loss of sexual reproduction and dwarfing in a small Metazoan. *PLOS ONE*, 5(9), e12854. <https://doi.org/10.1371/journal.pone.0012854>
- Thornhill, R., & Alcock, J. (2013). The Evolution of Insect Mating Systems. In *The Evolution of Insect Mating Systems*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674433960>
- Vakhrusheva, O. A., Mnatsakanova, E. A., Galimov, Y. R., Neretina, T. V., Gerasimov, E. S., Naumenko, S. A., Ozerova, S. G., Zalevsky, A. O., Yushenova, I. A., Rodriguez, F., Arkhipova, I. R., Penin, A. A., Logacheva, M. D., Bazykin, G. A., & Kondrashov, A. S. (2020). Genomic signatures of recombination in a natural population of the bdelloid rotifer *Adineta vaga*. *Nature Communications*, 11(1), 1. <https://doi.org/10.1038/s41467-020-19614-y>
- van der Kooij, C. J., Matthey-Doret, C., & Schwander, T. (2017). Evolution and comparative ecology of parthenogenesis in haplodiploid arthropods. *Evolution Letters*, 1(6), 304–316. <https://doi.org/10.1002/evl3.30>
- Veltsos, P., Fang, Y., Cossins, A. R., Snook, R. R., & Ritchie, M. G. (2017). Mating system manipulation and the evolution of sex-biased gene expression in *Drosophila*. *Nature Communications*, 8(1), 1. <https://doi.org/10.1038/s41467-017-02232-6>
- 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(12), 3493–3504. <https://doi.org/10.1046/j.1365-294X.2003.01998.x>
- Werren, J. H., Baldo, L., & Clark, M. E. (2008). *Wolbachia*: Master manipulators of invertebrate biology. *Nature Reviews Microbiology*, 6(10), 10. <https://doi.org/10.1038/nrmicro1969>
- White, M. J. D. (1977). *Animal cytology and evolution* (CUP archive).
- Wright, S., & Finnegan, D. (2001). Genome evolution: Sex and the transposable element. *Current Biology*, 11(8), R296–R299. [https://doi.org/10.1016/S0960-9822\(01\)00168-3](https://doi.org/10.1016/S0960-9822(01)00168-3)
- Xu, L.-Y., Wu, W.-T., Bi, N., Yan, Z.-J., Yang, F., Yang, W.-J., & Yang, J.-S. (2023). A cytological revisit on parthenogenetic *Artemia* and the deficiency of a meiosis-specific recombinase DMC1 in the possible transition from bisexuality to parthenogenesis. *Chromosoma*, 132(2), 89–103. <https://doi.org/10.1007/s00412-023-00790-x>
- Xu, S., Spitze, K., Ackerman, M. S., Ye, Z., Bright, L., Keith, N., Jackson, C. E., Shaw, J. R., & Lynch, M. (2015). Hybridization and the origin of contagious asexuality in *Daphnia pulex*. *Molecular Biology and Evolution*, 32(12), 3215–3225. <https://doi.org/10.1093/molbev/msv190>
- Yagound, B., Dogantzis, K. A., Zayed, A., Lim, J., Broekhuysse, P., Remnant, E. J., Beekman, M., Allsopp, M. H., Aamidor, S. E., Dim, O., Buchmann, G., & Oldroyd, B. P. (2020). A single gene causes thelytokous parthenogenesis, the defining feature of the Cape Honeybee *Apis mellifera capensis*. *Current Biology*, 30(12), 2248–2259.e6. <https://doi.org/10.1016/j.cub.2020.04.033>
- Ye, Z., Jiang, X., Pfrender, M. E., & Lynch, M. (2021). Genome-wide allele-specific expression in obligately asexual *Daphnia pulex* and the implications for the genetic basis of asexuality. *Genome Biology and Evolution*, 13(11), evab243. <https://doi.org/10.1093/gbe/evab243>

Chapter 1

In this first chapter, we characterized the candidate region associated with sex loss in one host race of the pea aphid with genome-wide association approaches. This manuscript has been published in *Genome Biology and Evolution*. I contributed to this study by doing bibliographic research for the introduction, analyzing the data using the BayPass program, and writing up the methods for this analysis.

Contrasting Evolutionary Patterns Between Sexual and Asexual Lineages in a Genomic Region Linked to Reproductive Mode Variation in the pea aphid

Maud Rimbault¹, Fabrice Legeai ^{1,2}, Jean Peccoud³, Lucie Mieuzet¹, Elsa Call ¹, Pierre Nouhaud ^{1,4}, H  l  ne Defendini¹, Fr  d  rique Mah  o¹, William Marande⁵, Nicolas Th  ron⁵, Denis Tagu¹, Ga  l Le Trionnaire¹, Jean-Christophe Simon¹, and Julie Jaqu  ry ^{1,*}

¹INRAE, UMR 1349, Institute of Genetics, Environment and Plant Protection, Le Rheu, France

²University of Rennes, Inria, CNRS, IRISA, Rennes, France

³Laboratoire Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, Unit   Mixte de Recherche 7267 Centre National de la Recherche Scientifique, Universit   de Poitiers, Poitiers CEDEX 9, France

⁴CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Univ Montpellier, Montpellier, France

⁵French Plant Genomic Resource Center, INRAE-CNRGV, Castanet Tolosan, France

*Corresponding author: E-mail: Julie.Jaquery@inrae.fr.

Accepted: September 12, 2023

Abstract

Although asexual lineages evolved from sexual lineages in many different taxa, the genetics of sex loss remains poorly understood. We addressed this issue in the pea aphid *Acyrtosiphon pisum*, whose natural populations encompass lineages performing cyclical parthenogenesis (CP) and producing one sexual generation per year, as well as obligate parthenogenetic (OP) lineages that can no longer produce sexual females but can still produce males. An SNP-based, whole-genome scan of CP and OP populations sequenced in pools (103 individuals from 6 populations) revealed that an X-linked region is associated with the variation in reproductive mode. This 840-kb region is highly divergent between CP and OP populations ($F_{ST} = 34.9\%$), with >2,000 SNPs or short Indels showing a high degree of association with the phenotypic trait. In OP populations specifically, this region also shows reduced diversity and Tajima's D , consistent with the OP phenotype being a derived trait in aphids. Interestingly, the low genetic differentiation between CP and OP populations at the rest of the genome ($F_{ST} = 2.5\%$) suggests gene flow between them. Males from OP lineages thus likely transmit their *op* allele to new genomic backgrounds. These genetic exchanges, combined with the selection of the OP and CP reproductive modes under different climates, probably contribute to the long-term persistence of the *cp* and *op* alleles.

Key words: life-cycle, reproductive polymorphism, sexual reproduction, asexuality, genome scan, cyclical parthenogenesis.

Significance

Asexual taxa occur in all major clades of eukaryotes and derive from sexual species. Yet, the genetic basis of these transitions is poorly understood because crosses cannot generally be performed to genetically map the ability to propagate asexually. As a result, a gene presumably responsible for sex loss has been identified in only one animal species—the Cape honeybee. Here, using pooled genome sequencing, we identified a 840-kb region (carrying 32 genes) that is associated with the transition to permanent asexuality in the pea aphid. We also revealed that sexual and asexual alleles likely diverged several hundred thousand years ago and that asexual lineages probably persist through contagious asexuality, whereby the few males they produce transmit asexual genes to sexual lineages. These results provide new insights into the mechanisms of coexistence of sexual and asexual lineages.

   The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

The prevalence of sexual reproduction in eukaryotes (Bell 1982) has long been considered as an evolutionary paradox, because sexual organisms transmit their genetic information twice less efficiently as asexual organisms do (Maynard Smith 1971). There is now a consensus that sex is favored over asexuality in the long term because it purges deleterious mutations that otherwise accumulate in asexual genomes, combines favorable mutations into genomes faster and generates genotypic diversity fueling adaptation (Muller 1964; Barton and Charlesworth 1998). Indeed, only few ancient asexual lineages exist (e.g., Mark Welch and Meselson 2000; Martens et al. 2003), indicating the inability of asexual lineages to persist over long evolutionary time due to long-term costs. However, how sex is maintained in the short term when sexual and asexual lineages coexist is still under debate (Hartfield and Keightley 2012). The loss of sexual reproduction is observed in many animal taxa such as squamates, fishes, insects, crustaceans, nematodes, and mollusks (Vrijenhoek et al. 1989; Schon et al. 2009). These frequent transitions from sexual to asexual reproduction reflect well the theoretical demographic advantage of asexual lineages over their sexual counterparts, which may allow them to persist over ecological times.

Sex may be lost by different ways (including interspecific hybridization, microorganism infection, spontaneous mutation, or spread of contagious asexuality elements) and at various frequency. The mechanisms of these losses affect the genetic features of the derived asexual lineages (Simon et al. 2003; van der Kooi and Schwander 2014). However, little is known about the genes underlying the shifts to asexuality. Indeed, one cannot use standard crossing techniques to genetically map the ability to propagate asexually (Neiman et al. 2014). Remarkably, certain species present lineages that have only partially lost sexual reproduction, allowing the identification of the genetic basis of sex loss using recombination-based approaches. Such crosses have revealed that the genetic mechanism responsible for the transitions from cyclical to obligate parthenogenesis in aphids (Dedryver et al. 2013; Jaquiéry et al. 2014), rotifers (Stelzer et al. 2010), and cladocerans (Lynch et al. 2008; Tucker et al. 2013; Xu et al. 2015), and from arrhenotoky to thelytoky in hymenopterans (Lattorff et al. 2005, 2007; Sandrock and Vorburger 2011; Aumer et al. 2017, 2019) involves only one or a few loci. However, in most cases, the precise location, the nature, and function of the genetic determinants of these shifts to obligate asexuality remain largely unknown.

The Cape honeybee is the animal species in which the gene responsible for sex loss is best characterized. Queens (and workers under certain conditions) produce haploid males via arrhenotokous parthenogenesis. Nevertheless, some workers

in the Cape honeybee have the ability to produce diploid eggs through thelytokous parthenogenesis. The extensive research into the genetic basis of this trait has yielded conflicting outcomes regarding the number of loci, the identification of the candidate gene, and the dominance/recessivity of the trait (Lattorff et al. 2005, 2007; Chapman et al. 2015; Wallberg et al. 2016; Aumer et al. 2017, 2019; Christmas et al. 2019; Yagound et al. 2020). The latest investigation points to a single gene (GB45239) that would be associated with thelytokous reproduction in workers (Yagound et al. 2020), the thelytokous allele being recessive. This gene encodes a protein that has structural similarity to SMC proteins, which typically play a role in chromosome assembly, segregation, and adhesion of sister chromatids. The functional characterization of the allelic variants at this candidate gene is crucial for establishing a causal link with thelytoky, given the controversy (Aumer et al. 2019; Christmas et al. 2019; Yagound et al. 2020).

Another well-studied system is *Daphnia pulex*, a crustacean reproducing by cyclical parthenogenesis, an alternation of many parthenogenetic generations and one sexual generation producing diapausing eggs, referred to as CP. In this species, sex-limited meiosis-suppressing genetic factors enable some lineages to produce diapausing eggs by parthenogenesis. These obligatory parthenogenetic lineages are called OP lineages. Genome sequencing of OP and CP lineages revealed that all *D. pulex* OP lineages share the same haplotypes in at least four genomic regions including almost two entire chromosomes and parts of two others (Lynch et al. 2008; Tucker et al. 2013; Xu et al. 2015), which have been acquired by hybridization with the close species *D. pullicaria*.

The identification of candidate loci for sex loss can also shed light on the origins and evolutionary dynamics of asexual lineages and/or asexual alleles. In the Cape honeybee, the allele associated with thelytoky appears to have emerged in this species and corresponds to a derived state (Yagound et al. 2020). In *D. pulex*, the large size of genomic regions associated with OP complicates the identification of candidate genes. However, some OP lineages still produce males, which can transmit the factors enabling permanent parthenogenetic reproduction when they mate with females from a CP lineage. These events create new OP lineages by so-called “contagious asexuality”. Analyses of rates of SNP conversion between OP and CP haplotypes within lineages revealed that all OP lineages of *D. pulex* were extremely young (22 years on average, Tucker et al. 2013). In contrast, the origin of the OP alleles is much older. Based on the synonymous divergence between the different OP haplotypes, it was estimated to have occurred between 1,250 and 187,000 years ago, corresponding to the divergence of the OP haplotypes clade from the homologous sequences in the exclusively sexual species *D. pullicaria* (Tucker et al. 2013). These results illustrate that, under contagious asexuality, the asexuality-conferring

allele can be markedly older than OP lineages themselves. Even though each OP lineage might be doomed to extinction, the ancient asexual allele can persist by spreading in new genomic backgrounds through males.

Aphids are another appropriate model for studying the genetic basis of the loss of sex. The ancestral mode of reproduction in this group is CP, but nearly 45% of the 5,000 aphid species have partially or completely lost sexual reproduction (Moran 1992). Typically, CP lineages undergo several successive generations of parthenogenesis (by viviparous parthenogenetic females) in spring and summer. In autumn, photoperiod shortening triggers the production of oviparous sexual females and males (Le Trionnaire et al. 2008). The winter-diapausing eggs resulting from sexual reproduction are the only frost-resistant stage of the aphid developmental cycle (Simon et al. 2002). They give birth to viviparous parthenogenetic females in the next spring, which start a new cycle.

Interestingly, some lineages have lost the ability to produce sexual females in response to the photoperiodic cues, and thus reproduce yearlong by viviparous parthenogenesis (Simon et al. 2002, 2010; Frantz et al. 2006). These OP lineages are demographically advantaged over CP lineages in mild winter regions, mainly because they do not go through a long egg diapause. However, they cannot survive in regions with harsh winters because they are unable to produce cold-resistant eggs (Moran 1992). Thus, selection by climate results in a geographical distribution of reproductive phenotypes where OP lineages occupy regions with mild winters and CP lineages those with cold winters (Defendini et al. 2023), both co-occurring in areas with intermediate or fluctuating climates (Rispe and Pierre 1998; Simon et al. 2002, 2010). Interestingly, many OP lineages have retained the capacity to produce males in autumn, so that gene flow between OP and CP lineages may occur in the wild (Halkett et al. 2008; Dedryver et al. 2013; Jaquiéry et al. 2014). In addition, since OP-produced males are usually fertile (Dedryver et al. 2019; Defendini et al. 2023), they can be crossed with CP females to identify the genetic basis of reproductive mode variation.

In the pea aphid *Acyrtosiphon pisum*, such crosses have revealed that the OP phenotype was recessive (Jaquiéry et al. 2014). The combination of two complementary approaches—QTL mapping and low-resolution genome scan using microsatellite markers on populations submitted to divergent selection for reproductive mode—pinpointed a 10-cM genomic region located on the X chromosome associated with this trait (Jaquiéry et al. 2014). However, none of the ~24,000 scaffolds constituting the ~540-Mb pea aphid genome sequence was anchored to any of the four chromosomes (IAGC 2010) and most of the scaffolds longer than 150 kb contained assembly errors associating unlinked chromosomal regions (Jaquiéry et al. 2018). As a result, the genomic context of microsatellites linked to

reproductive phenotypes could not be established. The recent release of an improved assembly of the pea aphid genome (Li et al. 2019), in which the four largest scaffolds correspond to the four chromosomes, provides an excellent opportunity to resolve this issue.

This study aims at finely characterizing the genomic region(s) associated with the variation of reproductive mode in the pea aphid and gaining functional and evolutionary insights into the genetic determinants of the loss of sex. To this end, we performed a high-resolution genome scan based on a pooled sequencing of 103 individuals from three OP and three CP populations. These individual samples had already been used in the previous low-resolution genome scan based on 439 microsatellite loci that identified one main candidate region associated with reproductive mode variation in *A. pisum* (Jaquiéry et al. 2014). The improved genome assembly combined with the millions of SNP markers scattered through the genome led to the identification of a major 840-kb genomic region showing strong genetic differentiation between OP and CP populations, and locating within the QTL locus previously identified by Jaquiéry et al. (2014). A thorough analysis of the variants present in this region was performed in an attempt to narrow down the list of candidate genes underlying the variation in reproductive mode in the pea aphid and to approximate the divergence time between the *op* and *cp* alleles.

Results

Genetic Structure of OP and CP Populations

A total of three OP populations, each consisting of 14 genetically distinct *A. pisum* lineages, and three CP populations, each composed of 20–21 lineages, were collected in alfalfa fields in Eastern Europe as described in Jaquiéry et al. (2014). These six populations were sequenced in pool, with two replicates per population, resulting in a total of 12 Pool-seq libraries. The number of 100 bp Illumina pair-end reads obtained ranged from 12 to 16 million per library, and from 24 to 28 million per population (table 1). The depth of sequencing, which varied from 15.1 to 20.4 depending on the libraries, enabled a total of 11 million SNPs to be identified. The PCA analysis of these SNPs separated the three OP populations from the three CP populations on the first axis, while the second axis mainly distinguished the OP populations (supplementary file 1, Supplementary Material online). As expected, sequencing replicates of the same population grouped together. Pairwise F_{ST} analyses revealed no differentiation between pairs of populations of the same reproductive mode. However, there was a slight differentiation between reproductive modes, with an average F_{ST} of 0.009 (see supplementary file 2, Supplementary Material online for

Table 1Geographical Origins of the *Acyrtosiphon pisum* Populations Collected on *Medicago sativa*

Reproductive mode	Location	Pop ID	Latitude/ Longitude	Number of lineages per pool	Library ID	Number of sequenced pairs	Number of mapped pairs	Properly paired	Properly paired w/o dups	Mean depth per library	
CP	Saint-Prex— Switzerland	SI	46°28' N 6°26' E	21	SI02	156,636,276	152,675,721 (97.5%)	144,341,466 (92.2%)	140,910,102 (90.0%)	20.4	
					SI08	124,231,404	120,516,514 (97.0)	114,359,174 (92.1%)	112,081,028 (90.2%)	16.2	
	Ranspach— France	VI	48°01' N 7°33' E	20	VI03	150,578,562	143,861,689 (95.5%)	135,517,068 (90.0%)	132,775,984 (88.2%)	18.8	
					VI09	129,361,706	124,600,198 (96.3%)	117,371,582 (90.7%)	115,159,748 (89.0%)	16.2	
	Mirecourt— France	Mil	48°16' N 6°06' E	20	Mil01	130,335,444	128,553,839 (98.6%)	122,795,378 (94.21%)	120,219,908 (92.2%)	120,219,908 (92.2%)	17.3
					Mil07	139,252,574	134,680,780 (96.7%)	127,537,064 (91.6%)	124,638,190 (89.5%)	17.9	
	OP	Castelnaudary —France	Cast	43°19' N 1°57' E	14	Cast04	144,133,842	138,947,572 (96.4%)	131,604,542 (91.3%)	129,063,056 (89.5%)	18.8
Cast10						121,695,798	115,774,856 (95.1%)	109,214,406 (89.7%)	107,442,998 (88.3%)	15.6	
Gers—France		Gers	43°57' N 0°22' E	14	Gers05	121,090,996	118,075,107 (97.5%)	110,620,444 (91.4%)	108,617,158 (89.7%)	15.8	
					Gers11	117,002,074	113,374,313 (96.9%)	106,802,538 (91.3%)	104,951,584 (89.7%)	15.1	
Lusignan— France		Lus	46°24' N 0°04' E	14	Lus06	122,039,296	117,497,862 (96.3%)	110,935,788 (90.9%)	109,002,292 (89.3%)	109,002,292 (89.3%)	15.8
					Lus12	133,230,226	129,092,091 (96.9%)	121,814,874 (91.4%)	119,566,626 (89.7%)	17.2	

These samples have already been used in the genome scan carried out in Jaquiéry et al. (2014). The reproductive mode (CP for cyclical parthenogenesis and OP for obligate parthenogenesis) as well as information on Illumina sequencing and read mapping are shown (values in brackets represent the percentage of the total number of pairs sequenced).

more detailed information per population). When populations were grouped by reproductive mode (enabling a more precise estimate of allelic frequencies in each reproductive mode), the genome-wide F_{ST} reached 0.025 between the two reproductive modes.

A Major Genomic Region is Associated With Reproductive Mode Variation

To identify regions possibly associated with reproductive mode variation, we investigated whether some genomic regions showed unusually high F_{ST} between OP and CP populations. We also ran BayPass (Olazcuaga et al. 2020), which aims at identifying SNPs associated with a binary trait (here OP or CP reproductive mode) while considering population structure, using the $C2$ statistic. Visual inspection of the $C2$ statistic and 20-kb sliding windows of F_{ST} along chromosomes revealed two genomic regions with high F_{ST} and $C2$ values (supplementary file 3, Supplementary Material online): a very short one (~30 kb) on chromosome 1 and a larger one on the X chromosome. However, the short region on chromosome 1 was found to be misplaced in the v3.0 reference genome (Li et al. 2019) and actually locates

on the X chromosome 2 Mb away from the region of highest F_{ST} (supplementary files 4 and 5, Supplementary Material online). Figure 1 therefore shows F_{ST} and the $C2$ statistic along a corrected genome in which the misplaced region has been moved to its true position. All genomic regions with mean F_{ST} values higher than 0.25 colocalized in the middle of the X chromosome (see fig. 1A and B). At least six secondary regions present F_{ST} values comprised between 0.2 and 0.25. The $C2$ statistic also provided a strong support for the main peak in the middle of the X chromosome. Of the 871 SNPs with a $C2$ value above 60 (corresponding to a highly significant association, $P < 10^{-14}$), 867 were located in the middle of the X chromosome. The remaining four SNPs were located in three different regions, none of which overlapped with the six secondary regions of moderate F_{ST} . The lack of agreement between F_{ST} and $C2$ methods and the low number of SNPs involved make these nine regions much less reliable than the main outlier region regarding their association with reproductive mode.

Focusing on the outlier region in the middle of the X chromosome, we observed particularly high F_{ST} and $C2$

values for a ~840-kb region (highlighted in blue in fig. 1B and D). This region is flanked by others showing slightly higher values than the rest of the genome, suggesting hitchhiking effect. However, none of these flanking regions was supported by both approaches: for example, the region near position 56 Mb (fig. 1B) contains an unusually low number of SNPs (see [supplementary file 6, Supplementary Material](#) online), none of which showed significant association with the reproductive mode according to the BayPass C2 statistic. Based on these results, an 840-kb region from position 62,895,000 to 63,735,000 on the X chromosome, that contained many windows with F_{ST} above 0.4 and 794 out of 871 SNPs with a C2 value above 60, was arbitrarily delimited as the main candidate region (fig. 1A and C). Importantly, this region locates at only 750 kb from the microsatellite marker having the strongest association with reproductive mode variation in QTLs and genome scan analyses using a low density of markers (Jaquiéry et al. 2014). This 840-kb region is highly divergent between CP and OP populations (average F_{ST} = 0.349) and contains 1,843 SNPs and 240 indels with F_{ST} > 0.5 between OP and CP populations, a value that denotes very different allelic frequencies between these two population types. This 840-kb region also showed elevated differentiation in every pair of populations differing in their reproductive mode, whereas no such pattern appeared for any pair of populations with the same reproductive mode (fig. 2). No other region showed consistent high differentiation between pairs of populations with different reproductive modes (fig. 2), further supporting a main region associated with reproductive mode variation.

To investigate the selection regimes acting on OP and CP populations, heterozygosity, Tajima's D and Fay and Wu's H (F_{WH}) were measured per reproductive mode using 100-kb windows along the genome (fig. 3). Genome-wide median heterozygosities were close in the OP and CP populations ($H_{E\ OP} = 0.280$ and $H_{E\ CP} = 0.276$) though they differed significantly ($W = 6,623,500$, $P < 10^{-15}$, two-sided Wilcoxon test). Heterozygosity in the 840-kb candidate region was significantly reduced compared with the rest of the genome in OP populations ($H_{E\ OP\ out} = 0.14$, $U = 15$, $P = 1.01 \times 10^{-06}$, two-sided Mann–Whitney test, fig. 3F), lying within the lower 0.06th percentile of the distribution. Contrastingly, heterozygosity in CP populations ($H_{E\ CP\ out} = 0.293$) was significantly higher than genome-wide median heterozygosity ($U = 27,589$, $P = 0.018$, fig. 3I). Other X-linked regions with moderately high F_{ST} values (from 0.1 to 0.15) showed reduced diversity in CP but not in OP populations (fig. 3G). Genome-wide median Tajima's D were significantly higher in OP than in CP populations (Tajima's $D_{OP} = -0.208$, Tajima's $D_{CP} = -0.292$, $W = 8,578,500$, $P < 10^{-15}$, two-sided Wilcoxon test). The 840-kb region also stood out in OP populations as being characterized by Tajima's D values (Tajima's $D_{OP\ out} = -0.890$) being in the lower 0.8th percentile of the

genome distribution. These values are significantly lower than those measured in the rest of the genome ($U = 503$, $P < 10^{-05}$, two-sided Mann–Whitney test, fig. 3L). No reduction in Tajima's D values was observed in the candidate region in CP populations compared to the rest of the genome (Tajima's $D_{CP\ out} = -0.287$, $U = 19,685$, $P = 0.75$, two-sided Mann–Whitney test, fig. 3O). To characterize the selective regimes that may have reduced the genetic diversity in OP populations, we measured F_{WH} . We observed no reduction of F_{WH} in the candidate region in OP populations, the values being actually significantly higher than the genome-wide median ($F_{WH\ OP\ out} = -0.122$, $F_{WH\ OP} = -0.570$, $U = 33,484$, $P < 10^{-4}$, fig. 3R), providing no support for a hard sweep (Garud et al. 2021). No difference was observed in CP populations ($F_{WH\ CP\ out} = -0.458$ and $F_{WH\ CP} = -0.508$, $U = 20,187$, $P = 0.67$, fig. 3U). These patterns suggest selection against lineages not carrying the *op* allele(s) in mild winter regions, leading to high F_{ST} between reproductive modes and reduced H_E and Tajima's D in OP only in the candidate region. However, the absence of negative F_{WH} does not suggest a rapid selective sweep of the *op* allele(s). Interestingly, a 200-kb region located at the left of the 840-kb candidate region also stood out (fig. 3, see also [supplementary file 7, Supplementary Material](#) online). It however showed similar patterns in OP and CP populations, including low Tajima's D and H_E , high F_{WH} , and low-to-moderate F_{ST} . We could then assign this region to a previously identified locus associated with wing polymorphism in males of the pea aphid (Braendle et al. 2005; Li et al. 2020), which may be subject to different selective pressures.

To better characterize the genetic variation linked to reproductive mode, we investigated the structure of the 840-kb candidate region in the OP and CP genomes. For this, we assembled the genomes of two clones (the OP X6-2 and the CP LSR1 lineages, see [supplementary file 4, Supplementary Material](#) online for assembly quality metrics) from long-read sequences. The candidate region was assembled in a single contig in both clones (fig. 4A, [supplementary file 5, Supplementary Material](#) online) and did not show any large structural rearrangement between these two individual genomes. The sequencing depth ratio OP/(OP + CP) computed over 2-kb windows from Pool-seq data (fig. 4B) also failed to reveal any large deletion in OP populations.

Age of Divergence of the *op* and *cp* Alleles

To approximate the age of the divergence of the *op* and *cp* alleles, we used three different approaches, as none of them was free of bias. The Nei and Li (1979) method based on absolute divergence (D_a , see eq. 1) and the experimentally estimated mutation rate in *A. pisum* (Fazalova and Nevado 2020) indicated a divergence time estimate of 183,129 years (95% CI: 130,513–281,076). The second approach relied on the number of substitutions between *op* and *cp*

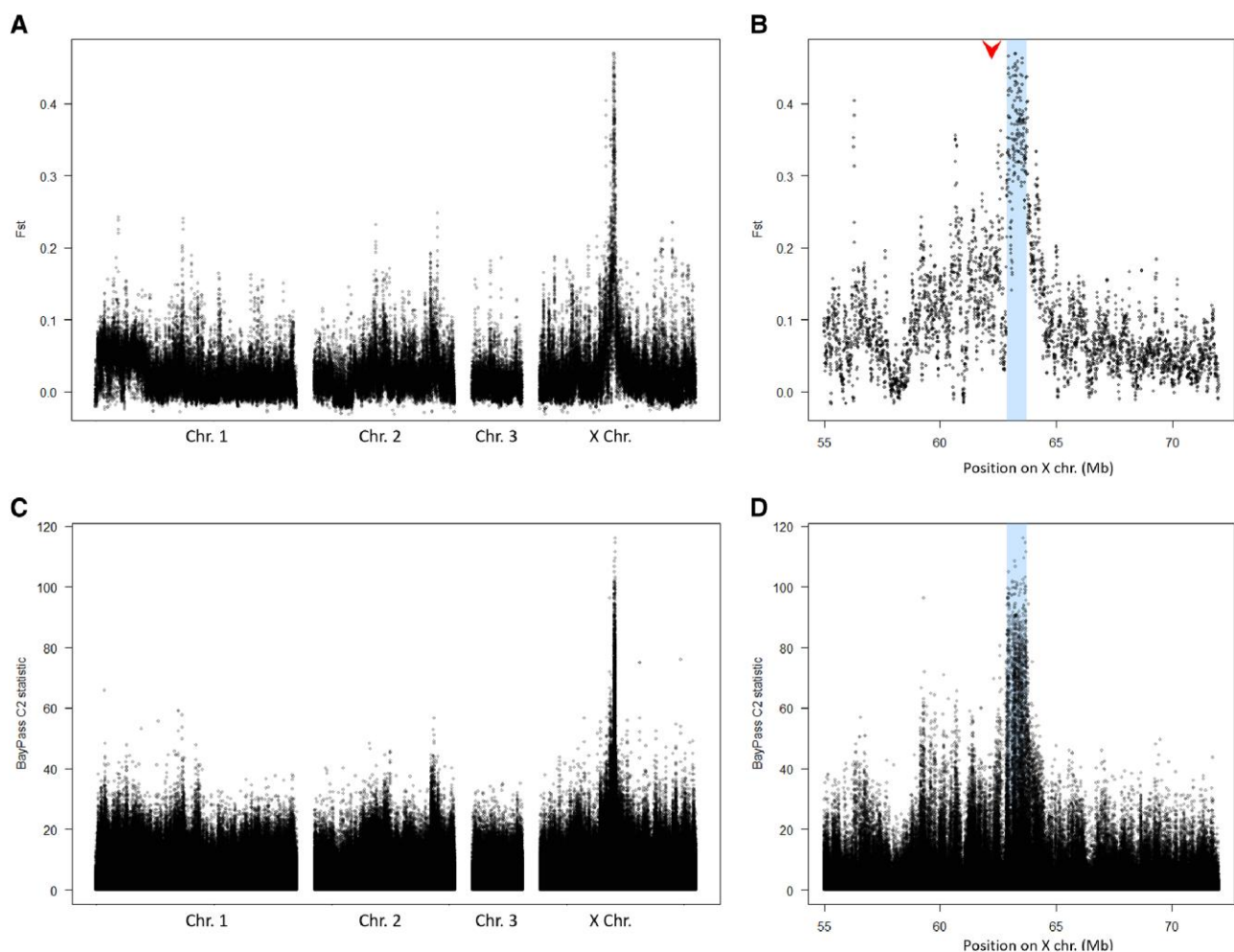


Fig. 1.—Detection of genomic regions associated with reproductive mode variation in *Acyrthosiphon pisum*. (A) Genetic differentiation (F_{ST}) between OP and CP populations (20-kb windows sliding by 5-kb steps). (B) Detail of the region of the X chromosome that contains the main outlier region. The arrow corresponds to the position of the outlier markers (in F_{ST} scan and QTL analysis) identified in Jaquéry et al. (2014). (C) C2 statistic from BayPass for each SNP for the whole genome and (D) for the main outlier region. The 840-kb region associated with reproductive mode variation is shown in blue.

consensus sequences reconstructed from the Pool-seq data ($N_{\text{mutated sites}}$) and the same estimated mutation rate (eq. 2). $N_{\text{mutated sites}}$ was 4,442 (95% CI: 4,340–4,530), such that the divergence time of the *cp* and *op* alleles was estimated to be 555,719 years (95% CI: 388,820–873,824). The third method relied on synonymous divergence (dS) calculated between *op* and *cp* alleles at the 32 genes in the candidate region based on the Pool-seq data. The dS was 0.00531 (95% CI: 0.00378–0.00676) and would result in a divergence time estimate of 515,437 years (95% CI: 366,734–656,212) based on the calibrated dS between *A. pisum* and *Myzus persicae* (Johnson et al. 2018; Mathers et al. 2020). Similar estimates were obtained when dS was measured between the resequenced genomes of an OP (LL01) and a CP individual (L9Ms03) ($dS = 0.0057$, 95% CI: 0.0049–0.0068, $T = 556,029$, 95% CI: 476,440–656,923).

Gene Content of the Candidate Region

The 840-kb candidate region associated with reproductive mode variation contains 32 predicted genes (table 2). Ten of these showed no homology with *Drosophila* proteins, nine of which were annotated as uncharacterized protein on NCBI, and one (LOC100159148) had homologies with a nuclear pore complex protein from *Salmo trutta* (table 2 and supplementary file 8, Supplementary Material online). The remaining 22 genes have *Drosophila* homologs, including seven that encode proteins of unknown function and 15 that are homologous to *Drosophila* genes with functional annotations and phenotypic characterizations. Interestingly, the amino acid sequences of these 15 genes all share the typical conserved protein domains identified in *Drosophila*, thus giving strong confidence in their annotation (supplementary file 8, Supplementary Material online). More precisely, four

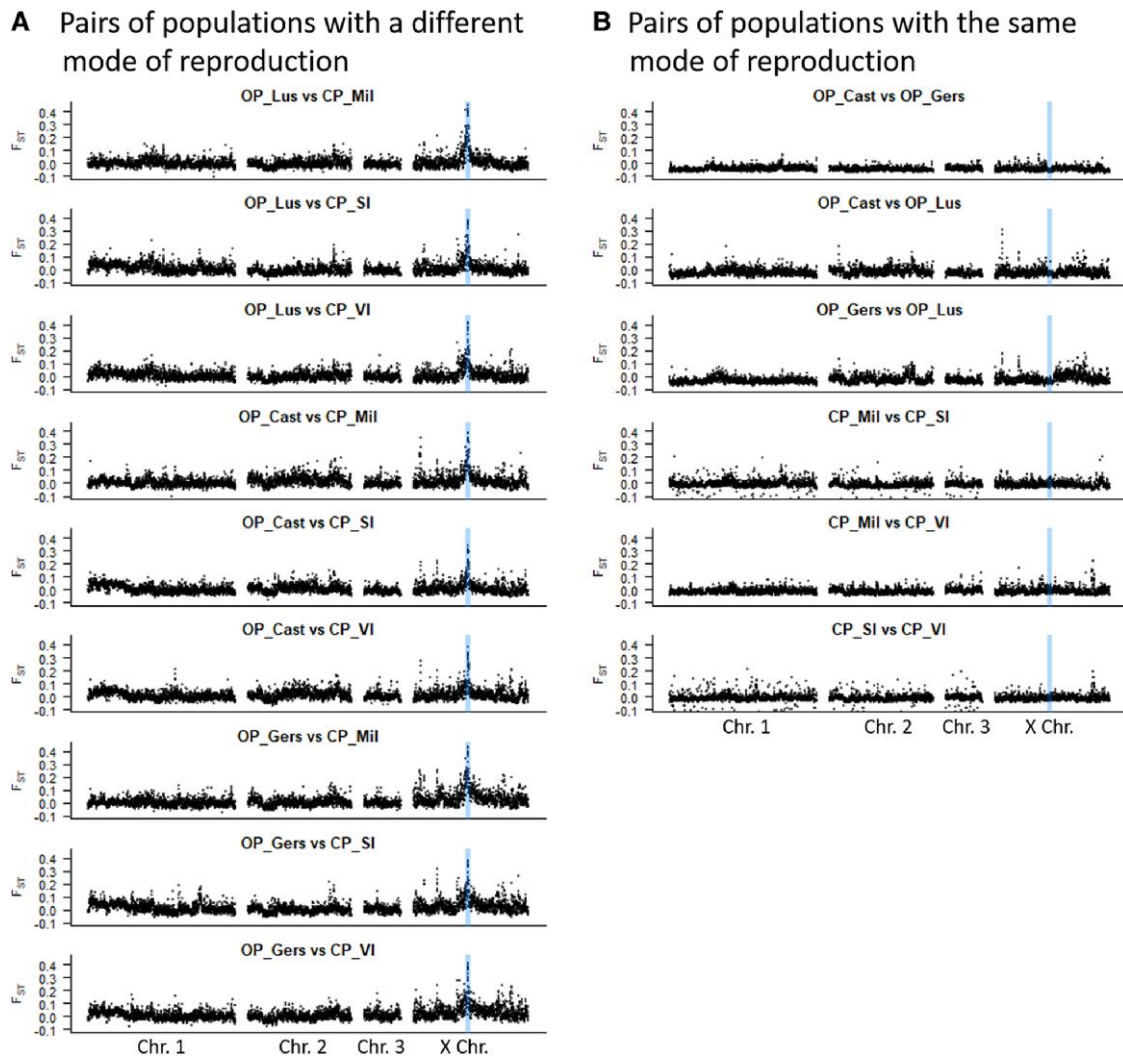


FIG. 2.— F_{ST} between pairs of populations in nonoverlapping 100-kb windows sliding along the genome of *Acyrtosiphon pisum*. (A) Pairwise comparisons between OP and CP populations. (B) Pairwise comparisons between populations with the same reproductive mode. The 840-kb candidate region is shown in blue.

are annotated as transcription factors, three of them sharing typical features of zinc-finger proteins (LOC100159233, LOC100161275, LOC107882169). Seven genes are homologous to genes coding for enzymes known to be involved in general metabolism in *Drosophila*: a trimethylguanosine synthase (LOC100570687), a sphingomyelin phosphodiesterase (LOC100169137), an N-acetylglucosaminyltransferase (LOC100569179), a protein kinase (LOC100161186), a fatty acyl-coA reductase (LOC100169017), a Rho GTPase activating protein (LOC100163133), and a cysteine-type peptidase (LOC100163837). Finally, the four remaining genes are homologous to *Drosophila* genes for which phenotypic analyses of mutants revealed their involvement in key biological processes associated with germline and embryo development, including miRNA processing and RNA interference for *Cpb20* (LOC100570523) and *pasha* (LOC100168027), cell cycle

control for *APC10* (LOC100165999), and dopamine signaling for *punch* (LOC100164133).

Among variants of the candidate region that showed large differences in allele frequencies between OP and CP populations ($F_{ST} > 0.5$), 38 impacted protein sequences (table 2 and supplementary file 8, Supplementary Material online). These included 35 missense variants, 1 frameshift variant, 1 conservative in-frame indel, and 1 nonsense variant (table 2 and supplementary file 8, Supplementary Material online), affecting a total of 11 genes. Five of these are homologous to genes encoding uncharacterized proteins. Three genes with homologs in *Drosophila*—*Cbp20* (LOC100570523), *Fatty acyl-CoA reductase* (LOC100169017), and *RhoGAP102A* (LOC100163133)—display one or two nonsynonymous SNPs outside the typical functional domains of these

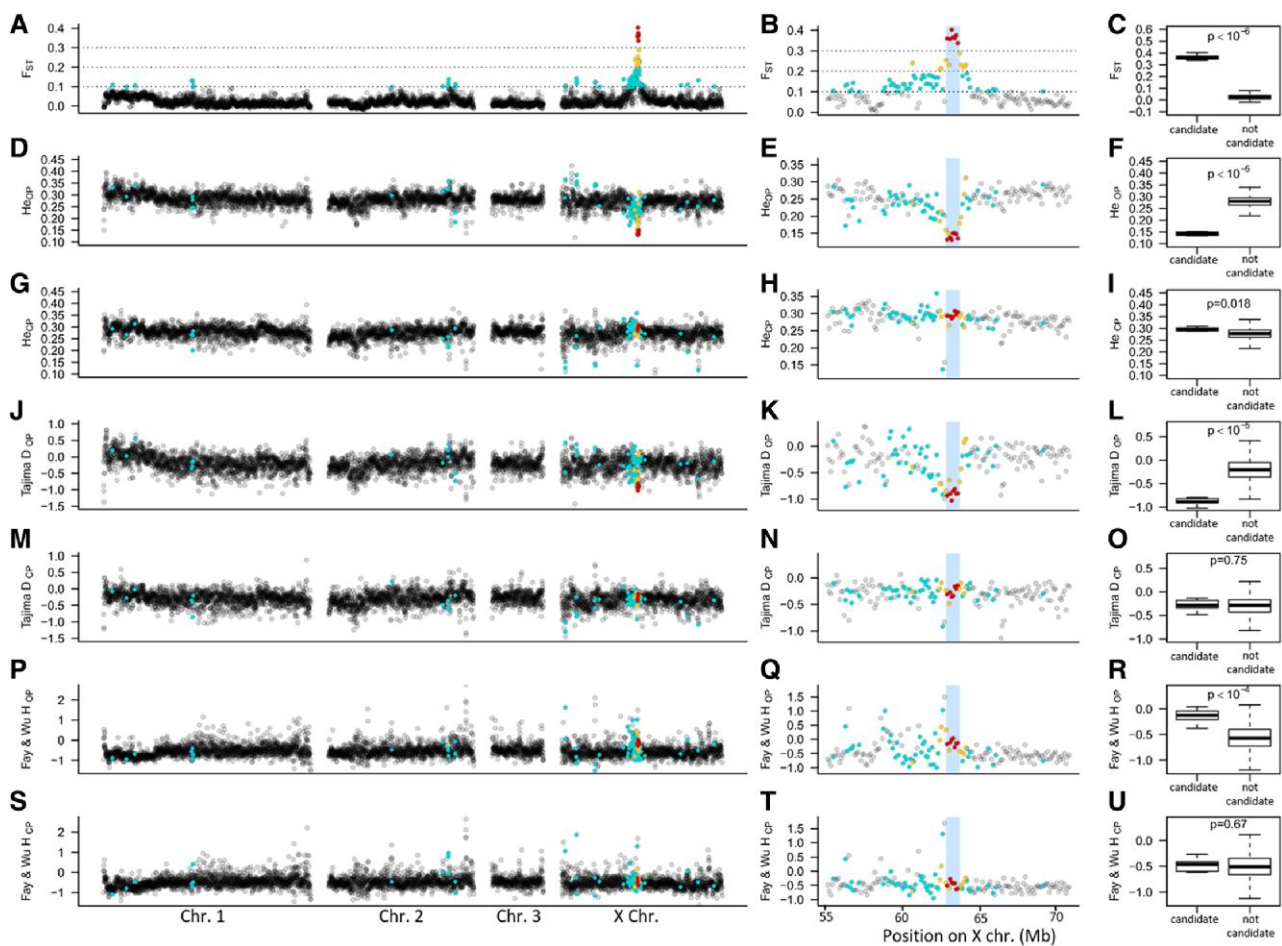


FIG. 3.—Population genetic indices calculated along the chromosomes of *Acyrthosiphon pisum* in nonoverlapping 100-kb windows. The plots in the first column (panels A, D, G, J, M, P, and S) show these indices for the whole genome, those in the second column for the part of the X chromosome carrying the main outlier region (highlighted in blue), and the boxplots in the third column compare the value of the eight 100-kb windows contained within the 840-kb candidate region (“candidate”) with that for the rest of the genome (“not candidate”). *P*-values on boxplots were obtained with two-sided Mann–Whitney tests. Panels A, B, C: F_{ST} between OP and CP populations; panels D, E, F and G, H, I show expected heterozygosity in OP and CP populations, respectively; panels J, K, L and M, N, O show Tajima’s *D* in OP and CP populations, respectively, and panels P, Q, R and S, T, U show Fay and Wu’s *H* for OP and CP populations, respectively. Each point (a 100-kb window) was colored according to its average F_{ST} value (red for windows with an average F_{ST} above 0.3, yellow for F_{ST} between 0.2 and 0.3, blue for those with F_{ST} between 0.1 and 0.2, gray for those below 0.1) in order to facilitate the visualization of the possible relationships between regions with high F_{ST} values and their values at other indices.

proteins. Interestingly, LOC100169137—homologous to a sphingomyelin phosphodiesterase—displays two SNPs in its mit_SMPDase domain, both changing the chemical property of the corresponding amino acid. Finally, two genes sharing features of zinc-finger transcription factors (LOC100159233 and LOC107882169) show polymorphism possibly resulting in truncated proteins in OP lineages. The remaining 21 genes of the region do not display any polymorphism changing the protein sequence between OP and CP lineages.

There was no clear evidence for large indels associated with reproductive mode variation within the 32 genes of the candidate region, as most (29) showed similar

sequencing depth in OP and CP populations (table 2 and supplementary file 9, Supplementary Material online). For each of the five genes that had less than 90% of their length sufficiently sequenced in OP and CP populations (LOC100163229, LOC107883347, LOC100159148, LOC100167415, and LOC100159233; table 2), the same gene segment shows reduced sequencing depth in both types of populations (supplementary file 9, Supplementary Material online). For three genes (LOC100160994, LOC100159717, and LOC100573568), the percentage of gene length with sufficient sequencing depth was lower in OP than in CP populations. However, this difference was supported by only one

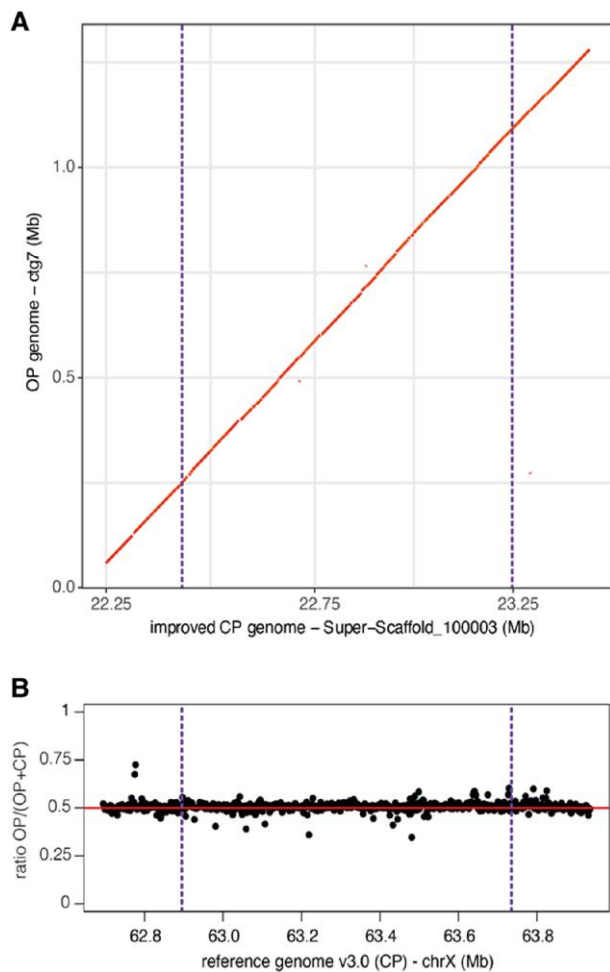


Fig. 4.—Structure of the 840-kb candidate region in OP and CP genome assemblies. (A) MUMmer alignment plot comparing parts of the two contigs (one per genome) containing the candidate region. (B) Normalized sequencing depth ratio $OP/(OP + CP)$ calculated over 2-kb windows along the candidate region. In both panels, the vertical dashed lines delimit the 840-kb candidate region.

population (out of six) in which the sequencing depth did not meet our criteria, hence failing to indicate consistent lack of coverage in all OP populations.

Discussion

In this study, we took advantage of a newly available genome assembled at the chromosomal level (Li et al. 2019) to precisely analyze the genomic differentiation between CP and OP populations of the pea aphid, enabling us to pinpoint a main genomic region associated with reproductive mode variation. This 840-kb candidate region carries 32 predicted genes and contains more than 2,000 SNPs and short indels that show strong differences in allelic frequencies between OP and CP populations, making it difficult to identify a causal gene. Population genetic indices revealed a

reduction in diversity in this region of 840 kb only in the OP populations, suggesting that selective events have affected only these populations.

A Major Region is Associated With Reproductive Polymorphism

The main candidate region associated with reproductive variation was relatively large (840 kb) and contained numerous highly differentiated SNPs throughout. We found little evidence for additional regions contributing to this trait. Indeed, other regions show much weaker association with reproductive mode (lower values of F_{ST} or $C2$) and none was supported by both F_{ST} and $C2$ statistics. Furthermore, our previous QTL analyses (Jaquiéry et al. 2014) showed a close association between reproductive phenotype and genotype at the microsatellite marker closest to the main 840-kb region identified here. We therefore focused on this main region, recognizing that secondary regions could explain the remaining proportion of variance not explained by the main QTL.

Population genetic indices revealed possible signatures of selection acting on the 840-kb region but only in the OP populations, which seems consistent with the OP phenotype being derived from the CP one (the ancestral state in aphids is CP, Davis 2012). Indeed, indices were highly asymmetric, heterozygosity, and Tajima's D being below the 1% of the lower extreme values in OP but close to the mean in CP populations. Interestingly, the reduced diversity at the candidate region in OP populations suggests that one or a few predominant *op* alleles are now present in wild populations. Diversity patterns are also consistent with the QTL-inferred dominance levels of alleles at this locus (Jaquiéry et al. 2014). Since *op* alleles are recessive, OP lineages are necessarily homozygous (explaining the low H_E in OP), whereas CP individuals can be either homozygous for the *cp* allele or heterozygous (hence the higher than average H_E in CP). Overall, these patterns suggest that strong environmental selection on this trait, which is also supported by the close association between reproductive phenotypes and winter climate in other pea aphid populations (Defendini et al. 2023). Nevertheless, *F_{WH}* provides no support for a hard selective sweep in OP populations. This statistic is powerful for detecting recent selective sweeps, but less effective for detecting soft, recurrent, or older sweeps (Kim and Stephan 2002; Przeworski 2002; Zeng et al. 2006), which could be the case here.

Origin of Reproductive Polymorphism

The large size of the region with high F_{ST} between reproductive modes is also intriguing. This pattern could result from a point mutation followed by hitchhiking of linked regions. As said above, a single ancient selective sweep, or repeated sweeps, could have occurred—two conditions in

Table 2

Annotation of the 32 Genes Predicted in the 840-kb Candidate Region Associated With Reproductive Mode in the pea aphid *Acyrthosiphon pisum*, and Numbers of Nonsynonymous Variants of Different Types Among Cyclical Parthenogenetic (CP) and Obligate Parthenogenetic (OP) Populations

Gene ID	NCBI gene description	<i>Drosophila</i> best hit	Annotation in <i>Drosophila</i>	Nonsense variants	Frameshift variants	Missense variants	Conservative in-frame deletions	% of positions with depth ≥ 20 in CP	% of positions with depth ≥ 20 in OP
LOC103308741	Uncharacterized protein	CG16854	Uncharacterized protein			1 ^a	1	90	90
LOC100570325	Uncharacterized protein	—	Uncharacterized protein			1		100	100
LOC100160994	Alpha-tocopherol transfer protein	CG10026	Uncharacterized protein					100	84
LOC100570523	Nuclear cap-binding protein subunit 2-like	<i>Cbp20</i>	Cap-binding protein			1		100	100
LOC100570687	Uncharacterized protein	<i>Tgs1</i>	Trimethylguanosine synthase 1					100	100
LOC100169137	Sphingomyelin phosphodiesterase 4	CG6962	Sphingomyelin phosphodiesterase			2 ^b		100	100
LOC100569418	Uncharacterized protein	—	Uncharacterized protein					100	100
LOC100569269	Uncharacterized protein	—	Uncharacterized protein					100	100
LOC100569179	UDP-N-acetylglucosamine	<i>sxc</i>	N-acetylglucosaminyltransferase					100	100
LOC103308943	Uncharacterized protein	—	Uncharacterized protein					100	100
LOC100163229	Putative nuclease HARBI1	CG43088	Uncharacterized protein					73	78
LOC107883347	Uncharacterized protein	CG4404	Uncharacterized protein					82	82
LOC100161186	MAPK/JAK/MRK overlapping kinase-like	CG42366	Mitogen-activated protein kinase					100	100
LOC100159148	Nuclear pore glycoprotein p62-like	—	Uncharacterized protein					89	87
LOC100168027	Microprocessor complex subunit DGCR8-like	<i>pasha</i>	Partner of drosha					100	99
LOC100165999	Anaphase-promoting complex subunit 10	<i>APC10</i>	Anaphase-promoting complex					100	100
LOC100568829	Uncharacterized protein	—	Uncharacterized protein			17		100	99
LOC100570789	Uncharacterized protein	—	Uncharacterized protein					100	100
LOC100568498	Uncharacterized protein	—	Uncharacterized protein					98	100
LOC100568585	Uncharacterized protein	—	Uncharacterized protein					100	100
LOC100168655	Scavenger receptor class B member 1	CG40006	Uncharacterized protein					100	100
LOC100169017	Fatty acyl-CoA reductase 1	CG1441	fatty acyl-CoA reductase 1			2		100	100
LOC100163133	Uncharacterized protein	<i>RhoGAP102A</i>	Rho GTPase activating protein			1		99	100
LOC100159717	Transcription factor glial cells missing-like	<i>gcm</i>	Glial cells missing					98	72
LOC100163837	Bleomycin hydrolase-like	CG1440	Cysteine-type peptidase					100	100
LOC100573568	uncharacterized protein	—	Uncharacterized protein			1		99	79

(continued)

Table 2 Continued

Gene ID	NCBI gene description	<i>Drosophila</i> best hit	Annotation in <i>Drosophila</i>	Nonsense variants	Frameshift variants	Missense variants	Conservative in-frame deletions	% of positions with depth ≥ 20 in CP	% of positions with depth ≥ 20 in OP
LOC100573386	Uncharacterized protein KIAA1841 homolog	CG6761	Uncharacterized protein					100	100
LOC100164133	GTP cyclohydrolase 1-like	<i>Punch</i>	GTP cyclohydrolase					100	100
LOC100167415	tigger transposable element-derived protein 4-like	<i>Cag</i>	Uncharacterized protein			2		43	42
LOC100159233	Zinc-finger protein 180-like	<i>CroI</i>	Zn finger protein	1		6		83	80
LOC100161275	Zinc-finger protein 271-like	<i>CroI</i>	Zn finger protein					100	100
LOC107882169	Zinc-finger protein 239-like	<i>Glass</i>	Zn finger protein		1	1		100	100

^aThe missense variant is localized in a DUF229 domain.

^bThe two missense variants are localized in a Mit_SMPDase domain.

which sweeps are not well detected by *FWH* (Kim and Stephan 2002; Przeworski 2002; Zeng et al. 2006). A reduction in recombination rate in the 840-kb region could also generate a uniform differentiation pattern over a large region. Such reduction may have resulted from a large inversion, a type of structural variation that is frequently involved in different polymorphisms (e.g., Joron et al. 2011; Mérot et al. 2020). However, our analyses based on de novo assemblies of OP and CP genomes do not support this hypothesis. Finally, high genetic differentiation between alleles may result from introgression from a divergent population. Indeed, the pea aphid is a complex of host-associated “races” that are still hybridizing (Peccoud et al. 2009, 2014, 2015). The allele that gives rise to an OP phenotype could already be present in another host race and simply have introgressed into the alfalfa race (the race studied here). Another possibility is that the introgression of a DNA segment from another host race (but not involved in the OP phenotype in that race) has caused an incompatibility in the alfalfa race. It could have disrupted the molecular cascade leading to the production of sexual females, resulting in an OP phenotype only in the alfalfa host race.

Whatever the mechanisms that led to the appearance of this polymorphism, the *op* and *cp* alleles appear to be quite divergent—on the order of 180,000–870,000 years. These estimates vary depending on the approach used and should be interpreted with caution as none of the dating methods used are free from bias. Two of them rely on an experimentally measured mutation rate (Fazalova and Nevado 2020) that did not include the sexual phase, yet meiosis is known to be mutagenic (Arbel-Eden and Simchen 2019). The D_a -based method further assumes no selection, and the dS -based method relies heavily on a calibrated estimate of interspecies divergence based on few fossils (Johnson et al. 2018). Interestingly, these estimates of allelic divergence time are almost framed by different estimations of the age of the radiation of the pea aphid complex, which also vary widely, from 18,000 to 47,000 years when using the divergence of the maternally inherited obligate endosymbiont *Buchnera aphidicola* in the host aphid lineages (Peccoud et al. 2009) to 419,000–772,000 years when using nuclear divergence (Fazalova and Nevado 2020). Such a large variation does not allow us to determine whether the *op* allele appeared before or after the radiation of the pea aphid complex. This uncertainty could be clarified by testing whether reproductive mode in other pea aphid host races (which also present OP and CP lineages, Frantz et al. 2006) is controlled by homologous *op* and *cp* alleles or some other genetic basis. However, hybridization between most pea aphid host races (Peccoud et al. 2009, 2014, 2015) may make it difficult to determine whether any shared polymorphism arose before the onset of their divergence.

In any case, the age of the *op* allele is likely to be much older than that of most OP lineages (i.e., clones) carrying

it. This is the case in *D. pulex*, where asexual lineages are estimated to be 22 years old on average, while the asexual allele is at least 1,250–187,000 years old (Tucker et al. 2013). In the pea aphid, the rare OP males can transmit their *op* allele if they succeed to mate with CP sexual females. This is most likely to occur in regions with intermediate winter temperature, where both types of lineages are expected to be sympatric. These crosses would allow the *op* allele to escape from linked deleterious mutations that may accumulate in OP lineages, and may also generate new OP lineages, ensuring the long-term persistence of OP populations (and *op* allele) through “contagious asexuality”. This mechanism would thus ensure the persistence of the *op* allele even in the face of recurrent extinctions of OP lineages, leading to higher age of divergence between *op* and *cp* alleles than the age of currently existing OP lineages. This scenario of contagious asexuality is supported by the low genetic differentiation between OP and CP populations (genome-wide average F_{ST} of 2.5%), which necessary requires some gene flow to homogenize genomes. Although contagious asexuality needs to be tested in nature for the pea aphid, previous works have verified this scenario in natural populations of other aphid species (Halkett et al. 2008). The strong contrast in winter temperatures between western European regions may further facilitate the long-term maintenance of the two alleles, allowing the two population types to persist stably in the areas to which they are adapted (e.g., Defendini et al. 2023).

Interestingly, another locus of considerable ecological importance locates at less than 150 kb of the 840-kb region associated with reproductive mode. This locus, called *aphicarus*, determines wing polymorphism in pea aphid males (Braendle et al. 2005; Li et al. 2020). We observed a specific genetic signature near *aphicarus* (fig. 3 and supplementary file 7, Supplementary Material online), which includes a drop of F_{ST} between OP and CP populations, a drop of both heterozygosity and Tajima's D in OP and CP populations and an increased F_{WH} in both populations. The frequency of the derived wingless allele is high in the alfalfa-adapted host race in Europe (around 90–95%, Frantz et al. 2009; Li et al. 2020), which could explain the low differentiation between OP and CP populations and the low H_E and Tajima's D at this locus. The high values for F_{WH} are puzzling, but could be related to balancing selection on wing polymorphism (the two phenotypes coexist in many host races, Li et al. 2019), or to the genotype of the outgroup (winged or wingless). In any case, we observed no association between *aphicarus* and the 840-kb candidate region, with reduction of diversities observed in both OP and CP populations for *aphicarus*, but limited to OP populations for reproductive mode variation. This pattern is compatible with the relative divergence time of the two polymorphisms, the wing polymorphism estimated to be at least six times older on the basis of dS (0.0386, Li et al. 2019) than the reproductive polymorphism (~ 0.0053 – 0.0057).

Functional Insights into the Genetic Determinants of the Loss of Sex

As our analyses did not reveal large structural variation between *cp* and *op* alleles, reproductive mode variation probably involves only small-sized polymorphisms. SNPs and small indels are frequent along the candidate region and may affect reproductive mode by altering the function of genes controlling the switch to the sexual phase. Previous transcriptomic studies have identified a number of genes that change expression during the transition from asexual to sexual reproduction in a CP lineage of the pea aphid (Le Trionnaire et al. 2007, 2008, 2012; Gallot et al. 2012). Interestingly, one of the 32 genes located within the candidate region (*cbp20*) corresponds to a gene differentially expressed between sexual and asexual germlines (Gallot et al. 2012). Its *Drosophila* homolog encodes an mRNA cap-binding protein involved in miRNA processing and gene silencing by RNAi and germline *Drosophila* mutants produce no eggs (Sabin et al. 2009). Remarkably, this gene also showed a nonsynonymous polymorphism with a high F_{ST} between the OP and CP populations. Although this variation lies outside the typical RNA recognition motif domain of the protein, functional analyses by CRISPR/Cas9 targeted mutagenesis are now required to assess whether this variation determines reproductive mode.

Ten other genes are affected by nonsynonymous polymorphisms between the OP and CP populations. Five are of unknown function and five have a predicted function with no apparent link to variation in reproductive mode. Among the latter are two genes containing zinc-finger domains that are truncated or show a frameshift, probably leading to nonfunctional proteins in OP lineages. However, these two proteins do not share strong similarities with well-characterized *Drosophila* transcription factors, making it difficult to predict the phenotypic consequences of their disruption.

Interestingly, three other genes of the candidate region (*Pasha*, *APC10*, and *punch*) share similarities with *Drosophila* genes whose functions could play a role in reproductive mode switch in aphids. *Pasha* is involved in miRNA biogenesis: germline mutants do not form cysts from the germarium and fail in oocyte fate determination (Azzam et al. 2012). *APC10* promotes metaphase to anaphase transition during the cell cycle, and germline mutants show defects in stem cells production (Liu et al. 2016). Finally, *punch* is involved in eye pigmentation and cell cycle control. Some mutants show defaults in dopamine synthesis and embryo development (Hsouna et al. 2007). None of these three genes contained nonsynonymous variants between the OP and CP populations, but polymorphisms outside protein-coding sequences could also control reproductive mode variation. Indeed, intergenic and intronic regions contain DNA motifs to which regulatory factors may bind. Transcriptomic analyses of OP and CP

lineages submitted to long and short photoperiod regimes would allow testing whether some of the genes of the candidate region are differentially expressed, thus whether they could control the reproductive polymorphism through differences in protein levels. A parallel can be drawn with *D. pulex*, where male production is genetically controlled (Innes and Dunbrack 1993; Innes 1997). A recent genome scan analysis pinpointed a single gene whose male-producing and nonmale-producing alleles differ by seven nonsynonymous substitutions (Ye et al. 2019). These alleles are also expressed at different levels in response to the environmental cue normally inducing the production of males (Ye et al. 2019). Whether pea aphid reproductive polymorphism is determined by expression levels, protein variants, or a combination of both remains an open question.

Conclusion

This work refines the size, location, and gene content of the locus associated with sex loss in the pea aphid. Further functional studies are needed to identify the gene(s) driving reproductive mode variation, and to determine whether variation in this trait depends on variation in protein sequence and/or protein levels. Transcriptomic analyses of OP and CP lineages exposed to long and short photoperiods should help to identify the causal gene(s) and underlying mechanisms. CRISPR/Cas9 targeted mutagenesis, which has been successfully developed in the pea aphid (Le Trionnaire et al. 2019), would then allow a functional validation of the role of candidate genes. Furthermore, exploring the genetic basis of sex loss in other host races and species should clarify whether reproductive mode variation, which is widespread in aphids (Moran 1992; Simon et al. 2002), relies on common or independent mechanisms, and whether introgression might be involved. Finally, sequencing individual OP lineages in various populations would allow assessing the accumulation of deleterious mutations in these clones and whether this factor primarily dictates their fate.

Materials and Methods

Aphid Sampling

This study is based on the *A. pisum* samples previously used to conduct a low-density microsatellite-based genome scan (Jaquiéry et al. 2014). Briefly, parthenogenetic females were collected on *Medicago sativa* in alfalfa-cultivated fields from six sampling sites in 2008 (table 1). Three sites located in north-east France and Switzerland, where only CP lineages can survive cold winters. The three other sites located in south-west France where winters are generally mild and therefore favor OP. For each of the six geographical populations, we succeeded to keep alive 14–21 genetically distinct clonal lineages, each initiated by a sampled

female (table 1). Clonal individuals for each lineage were collected in 2008 and stored in absolute ethanol at -20°C until DNA extraction in 2013.

Pool Sequencing

DNA was extracted from four fourth instar larvae per clonal lineage using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. After RNase treatment, DNA solutions were pooled in equimolar proportions for each population ensuring that clonal lineages contributed equivalent amounts of DNA to the pool. Two independent paired-end libraries were constructed per population from these DNA pools using the Genomic DNA Sample Preparation Kit (Illumina, San Diego, CA) (technical replicates). The resulting 12 libraries were sequenced in 2013 on four lanes of the Illumina HiSeq 2000 platform in a single 2×100 -cycle run using Illumina Sequencing Kit v3, producing between 11.7 and 15.7 million paired-end 100 bp reads per library (table 1). The raw data are publicly available at the Sequence Read Archive of the NCBI database, under the BioProject ID PRJNA454786.

Mapping

For read mapping, we used the v3.0 reference genome of the pea aphid (NCBI: *pea_aphid_22Mar2018_4r6ur*, Li et al. 2019). This assembly is 541 Mb in size and consists of four main scaffolds corresponding to the three autosomes and the X chromosome, and 21,915 additional short scaffolds not positioned on chromosomes (which account for 14% of the bases, Li et al. 2019). Paired-end reads were mapped to a fasta file containing the *A. pisum* reference genome v3.0 and the sequences of its known endosymbionts (Guyomar et al. 2018) with bwa-mem v0.7.10 (Li and Durbin 2009), using defaults parameters. Only primary and properly paired alignments were kept with SAMtools v1.6 (Li et al. 2009). Read pairs corresponding to duplicates were then identified with Picard MarkDuplicates v2.18.2 (<http://broadinstitute.github.io/picard/>) and removed. The final number of read pairs kept ranged from 10.5 to 14.1 million, resulting in a sequencing depth per library of 15.1–20.4x (see table 1 for mapping statistics).

Variant Calling

The 12 alignment (BAM) files corresponding to the 12 DNA libraries were merged in a single mpileup file using SAMtools (Li et al. 2009) and a sync file was created using Popoolation2 (Kofler et al. 2011) with default parameters except for a minimum base quality set to 20. Positions corresponding to the aphid symbiont and mitochondria genomes were removed from the sync file, to analyze the pea aphid nuclear genome only.

A total of 11,954,278 SNP positions were identified, in which the least frequent allele was represented by at least four reads ($MAC \geq 4$). This dataset also included 181,204 tri-allelic SNPs for which the third (least frequent) allele was represented by only one read. This third allele was ignored and the SNP was considered biallelic. We checked that this subdataset showed the same signal as the full dataset ([supplementary file 10, Supplementary Material online](#)).

To visualize the structure of the dataset and check that replicates from the same population cluster together as expected, a principal component analysis (PCA) was carried out on the 12 libraries with `prcomp` in R version 3.6.1 (R Core Team 2019). We used allele frequencies at 50,000 randomly drawn SNPs. Since the two libraries from each population grouped together ([supplementary file 2, Supplementary Material online](#)), we summed their allele counts as if they constituted only one library in all subsequent analyses. Thereafter, we refer to this dataset of 11,954,278 SNPs as the nonfiltered dataset.

Additional filters were then applied to select reliable and informative SNPs for the F_{ST} and heterozygosity estimation. First, only SNP positions with a sequencing depth higher than 20 and lower than 60 per population were considered, the mean depth ranging from 31 to 37 depending on the population. The upper limit of 60 was chosen to avoid duplicated genomic regions not resolved in the reference genome, and the lower limit of 20 to discard SNPs whose sampling was too low for reliable allele frequency estimates. Second, a minor allele frequency threshold of 5% was applied to eliminate SNPs harboring rare alleles and which are not informative for a F_{ST} -based genome scan. After applying these selection criteria, we obtained a filtered dataset of 4,633,747 SNPs.

Detection of Genomic Regions Associated With Reproductive Mode Variation

We ran BayPass under the core model using the nonfiltered SNP dataset to compute the $C2$ genetic differentiation statistic (Olazcuaga et al. 2020). This $C2$ statistic compares the standardized population allele frequencies (i.e., the allele frequencies corrected for the population structure) between the two groups of populations specified by the binary covariable of interest, here the reproductive mode (CP or OP). As the number of SNPs was very large and computing time increases nonlinearly with the number of SNPs, we divided the full dataset into 100 subdatasets containing only 1 SNP every 100 SNPs along the genome as recommended by Gautier et al. (2018). This strategy allowed analyzing all available SNPs while limiting the influence of linkage disequilibrium. After checking for convergence with three independent runs of the first subdataset using options `-nval 500 -thin 200 -npilot 10 -pilotlength 300`

`-burnin 500` for the Markov chain Monte Carlo algorithm, we ran the analysis on all 100 subsets.

In parallel, to estimate genetic differentiation (F_{ST}) between populations with different reproductive modes, we summed allele counts in the filtered dataset (4.6 million SNPs) for the three CP populations on one hand and for the three OP populations on the other hand. These counts were used to calculate F_{ST} at each SNP between reproductive modes with the R package `poolstat`, which implements F_{ST} estimates for Pool-seq data (Hivert et al. 2018). We then calculated the average F_{ST} within 20-kb windows sliding by 5-kb steps to smooth its variation along the genome and precisely identify regions of high differentiation. F_{ST} within nonoverlapping 100-kb windows between the two reproductive mode and between all possible pairs of populations were computed in the same fashion.

Heterozygosity (H_E , following Nei 1973) was calculated per type of populations (OP or CP) at each SNP using allele frequencies from the filtered dataset. The mean H_E per population type was then computed in 100-kb contiguous windows. To detect potential selective sweeps, Tajima's D and F_{WH} were calculated for each population type. For Tajima's D , we used the pileup-formatted SNP files for the pool samples of each reproductive mode (i.e., one OP and one CP population) generated previously. We randomly subsampled the datasets as recommended to achieve a uniform depth using `PoPoolation 1.2.2` (Kofler et al. 2011), using the following parameters: `-target-coverage 30 -max-coverage 360 -min-qual 20`. Tajima's D was then calculated using `PoPoolation 1.2.2` over 100-kb nonoverlapping windows with the following parameters: `-min-count 2 -min-covered-fraction 0.5`. Then, F_{WH} was calculated with `npstat` (Ferretti et al. 2013) for each type of population (OP or CP) on 100-kb windows. Input files consisted of the pileup-formatted SNP files for the pool samples of each reproductive mode, and a fasta file of an outgroup (here, a cryptic species of the pea aphid complex adapted to feed on *Ononis spinosa*, Peccoud et al. 2009). For this, Illumina pair-end 100 bp whole-genome re-sequencing data of a CPA *pisum* individual of the *Ononis* host race used in a previous study (Guyomar et al. 2018) was retrieved from NCBI (Project PRJNA255937—SRX661218). After filtering the reads with `fastp v0.20`, reads were aligned with `bwa-mem v0.7.17` on the v3.0 reference genome of the pea aphid with default parameters, and duplicates were marked with `GATK MarkDuplicates v4.1.4.1`. SNP variants were called with `GATK HaplotypeCaller v4.1.1` with options `-heterozygosity 0.001 -heterozygosity-stdev 0.01 -sample-ploidy 2`. SNP variants with a $GQ > 20$ were then used to modify the v3.0 genome sequence to produce a fasta sequence for this outgroup. Positions homozygous for the alternative allele were replaced by the alternative allele, and for those identified as heterozygous, alleles were drawn at random, with equiprobability (using the R function `rbinom`). We used the following parameters to run `npstat`: `-l 100000 -mincov 4 -maxcov 180 -minqual 20 -nolowfreq 3`.

Differences in H_E , Tajima's D and F_{WH} between the OP and CP populations were evaluated with two-sided Wilcoxon tests (using 100-kb nonoverlapping windows as the statistical unit). Within each reproductive mode, two-sided Mann–Whitney tests were used to test for differences in these population genetic indices between candidate region(s) for reproductive mode variation and noncandidate regions. Note that the lack of independence between windows that are physically linked within the same candidate region may pose an issue for these tests.

Comparison of the Structure of the Candidate Region in CP and OP Genomes

The above analyses identified a genomic region as main candidate to explain the variation in reproductive mode. To compare the structure of the candidate region between CP and OP genomes, we assembled the genome of an OP lineage (clone X6-2, Jaquiéry et al. 2014), as the *A. pisum* reference genome v3.0 (Li et al. 2019) was assembled from a CP lineage (clone LSR1, IAGC 2010). Oxford Nanopore technology was used to obtain long-read sequences from the OP lineage and to build a de novo genome assembly (see [supplementary file 4, Supplementary Material](#) online for details). We also found that the *A. pisum* reference genome v3.0 (Li et al. 2019) contained some small assembly errors which could impact our results (see Results section and [supplementary files 4 and 5, Supplementary Material](#) online). We therefore constructed a new assembly for the LSR1-CP lineage (referred to as “improved CP genome” hereafter) with ONT- and PacBio-generated long reads and optical map data ([supplementary file 4, Supplementary Material](#) online). We then compared the structure of genomes assemblies at a 1.25-Mb region containing the 840-kb candidate region using MUMmer v3.22 (Kurtz et al. 2004). Pairwise alignments of the CP and OP genome sequences were assessed using NUCmer v3.07. Results were filtered using the delta-filter script to keep optimal correspondence with a minimum length of 1,000 bp and a minimum alignment identity of 90%, and were visualized using MUMmerplot v3.5 (Kurtz et al. 2004). Complementarily, to investigate the possible deletion of short genomic regions in OP populations, we plotted the sequencing depth ratio OP/(OP + CP) from the Pool-seq data on the v3.0 genome sequence. The sequencing depths of the OP and CP populations were normalized prior to ratio calculation, so that a ratio of 0.5 is expected for genome segments presenting the same copy number in the OP and CP populations. To visualize results, we computed the average of this ratio over 2-kb nonoverlapping windows on the candidate region.

Age of Divergence of *op* and *cp* Alleles

To roughly estimate the divergence time between the *op* and *cp* alleles of the main candidate region, we used three

different approaches, each with its own limitations. The first two used the complete DNA segment of the candidate region (coding and noncoding sequences), while the third one used coding sequences only. Divergence time (T) was first estimated as

$$T = D_a / (2 \times \mu), \quad (1)$$

where D_a is the absolute divergence at the candidate region and μ is the substitution rate (Nei and Li 1979). A related estimate of T was obtained as

$$T = N_{\text{mutated sites}} / (2 \times N_{\text{sites}} \times \mu), \quad (2)$$

where $N_{\text{mutated sites}}$ is the number of observed substitutions between the two alleles over all callable sites N_{sites} . The third approach relied on the dS between *op* and *cp* alleles at the 32 genes from the candidate region and the dS between two aphid species of known divergence time to infer T assuming proportionality.

For the first approach (eq. 1), D_a was estimated from the allele frequencies in the nonfiltered Pool-seq data, as in Cruickshank and Hahn (2014). A mutation accumulation experiment in the pea aphid estimated the mutation rate μ_{parth} to 2.7×10^{-10} (95% CI: 1.9×10^{-10} – 3.5×10^{-10}) per parthenogenetic generation (Fazalova and Nevado 2020). The annual mutation rate for an OP lineage was thus estimated as $\mu_{\text{op}} = N_{\text{gen}} \times \mu_{\text{parth}}$, N_{gen} being the number of generations per year (estimated to 15). For a CP lineage, we followed Fazalova and Nevado (2020) and estimated the mutation rate as $\mu_{\text{cp}} = (N_{\text{gen}} - 1) \times \mu_{\text{parth}} + \mu_{\text{sex}}$, where μ_{sex} (2.96×10^{-9} ; 95% CI: 1.52×10^{-9} – 4.99×10^{-9}) is the average mutation rate per sexual generation in insects (Keightley et al. 2014, 2015; Yang et al. 2015; Liu et al. 2017; Oppold and Pfenninger 2017) as there is no such estimate for aphids. The mutation rate to consider in equation (1) was thus 5.39×10^{-9} (95% CI: 3.52×10^{-9} – 7.57×10^{-9}).

For the second approach (eq. 2), in addition to the mutation rate (obtained above), we had to measure the number of substitutions between the two alleles, hence to reconstruct *op* and *cp* consensus sequences. For each reproductive mode, we randomly drawn alleles at polymorphic SNPs in the 840-kb region, where the probability of being drawn equals the allele frequency (using the R function `rbinom`). Then, we counted the number of substitutions between the *op* and *cp* consensus sequences (i.e. $N_{\text{mutated sites}}$). This step was repeated 100 times to get 95% confidence interval for $N_{\text{mutated sites}}$. N_{sites} was determined as the number of callable sites with sequencing depth ≥ 20 in every population (740,918).

For the third approach, coding sequences were extracted from each replicate of the consensus *op* and *cp* sequences generated above, and median dS between *op* and *cp* alleles at the 32 genes from the candidate region were computed with the `seqinr` R package (Charif and Lobry 2007). This

step was repeated on the 100 replicates of consensus sequences to get 95% CI for dS . Divergence time was then estimated by proportionality, assuming constant synonymous mutation rate between cp and op alleles and between the pea aphid and the peach–potato aphid *M. persicae*, whose divergence is estimated at some 22 million years ago and corresponds to a dS of 0.2268 (Johnson et al. 2018; Mathers et al. 2020). dS at the 32 genes were also estimated using the genomes of two individuals of the alfalfa host race, which were previously sequenced using 100 bp Illumina read pairs. One is of CP phenotype (clone L9Ms03—Project PRJNA255937—SRX661210, see Guyomar et al. 2018) and the other of OP phenotype (clone LL01—PRJNA255937—SRX20811676). Read filtering, mapping, SNPs calling and the construction of consensus sequences were carried out as previously described for the resequenced genome of an individual from the *O. spinosa* host race (see the nsstat analyses), and dS between op and cp alleles at the 32 genes computed with seqinR. A total of 100 replicates of the consensus sequences were built to calculate 95% CI, since these sequenced individuals were heterozygous at some SNPs.

Gene and Variant Annotation in the Main Candidate Region

Amino acid sequences of the predicted genes present in the 840-kb candidate region were retrieved from the v3.0 version of the pea aphid genome assembly. Annotations for these genes were obtained from the general feature format (gff) file available on NCBI (GCF_005508785.1_pea_aphid_22Mar2018_4r6ur_genomic.gff.gz). Whenever a gene had multiple predicted transcripts, we only kept the longest transcript. A BlastP analysis (Altschul et al. 1990) was then performed against Flybase (<http://flybase.org/>) to identify the closest *Drosophila* homolog for each of these aphid genes (at $P < 10^{-7}$). Conserved protein domains were identified and annotated for each gene using the SMART web resources (<http://smart.embl-heidelberg.de/>; Letunic et al. 2020) with the “normal” mode and a significance level of 10^{-10} . To detect potentially causal polymorphisms, we examined the variants (SNPs and short indels) from the candidate region for reproductive mode variation showing F_{ST} above 0.5 and a sequencing depth ≥ 20 in every population. These variants were classified according to their impact on gene structure by SnpEff v4.3t (Cingolani et al. 2012) with default parameters and using the GFF file available on NCBI. Variants with moderate-to-high predicted impact were retained for further analysis. This includes variants resulting in premature stop codons, frameshifts, missenses, or conservative in-frame indels.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

This work was supported by grants from the French Research Agency (SexAphid ANR-09-GENM-017-001, Speciaphid ANR-11-BSV7-0005 and Mecadapt ANR-18-CE02-0012), the INRAE-SPE Department (AAP GenAsex and half a PhD grant for H.D. and P.N.), Région Bretagne (ARED, half a PhD grant for H.D. and P.N.), and the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Action (grant agreement no. 764840 for the ITN IGNITE project). We thank Editor Aida Andrés and one referee for constructive comments on a previous draft of the manuscript.

Author Contribution

J.J. and J.C.S. conceived and designed the study. L.M. and F.M. carried out the DNA extractions and prepared the samples. F.L., W.M., J.P., D.T. and N.T. performed genome assembly. M.R., J.J., F.L., H.D., G.L.T., E.C. and P.N. performed the analyses. J.J., J.C.S., J.P., G.L.T., M.R., H.D., F.L. and W.M. wrote the manuscript. All authors reviewed and revised the manuscript.

Data Availability

Raw sequence reads are deposited on NCBI (PRJNA454786, PRJNA745262, PRJNA255937-samples SRX661210, SRX20811676, and SRX661218). Genome assemblies are available at the following permanent addresses: https://bipaa.genouest.org/sp/acyrthosiphon_pisum/download/genome/LSR1_CP/; https://bipaa.genouest.org/sp/acyrthosiphon_pisum/download/genome/OP/. The data and scripts are available on Zenodo: doi.org/10.5281/zenodo.8116727.

Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Arbel-Eden A, Simchen G. 2019. Elevated mutagenicity in meiosis and its mechanism. *Bioessays* 41:e1800235.
- Aumer D, et al. 2019. A single SNP turns a social honey bee (*Apis mellifera*) worker into a selfish parasite. *Mol Biol Evol.* 36:516–526.
- Aumer D, Allsopp MH, Lattorff HMG, Moritz RFA, Jarosch-Perlow A. 2017. Thelytoky in Cape honeybees (*Apis mellifera capensis*) is controlled by a single recessive locus. *Apidologie* 48:401–410.
- Azzam G, Smibert P, Lai EC, Liu J-L. 2012. *Drosophila* Argonaute 1 and its miRNA biogenesis partners are required for oocyte formation and germline cell division. *Dev Biol.* 365:384–394.
- Barton NH, Charlesworth B. 1998. Why sex and recombination? *Science* 281:1986–1990.
- Bell G. 1982. The masterpiece of nature: the evolution and genetics of sexuality. Berkeley: University of California Press.
- Braendle C, Caillaud MC, Stern DL. 2005. Genetic mapping of aphicarus—a sex-linked locus controlling a wing polymorphism in the pea aphid (*Acyrtosiphon pisum*). *Heredity* (Edinb) 94:435–442.

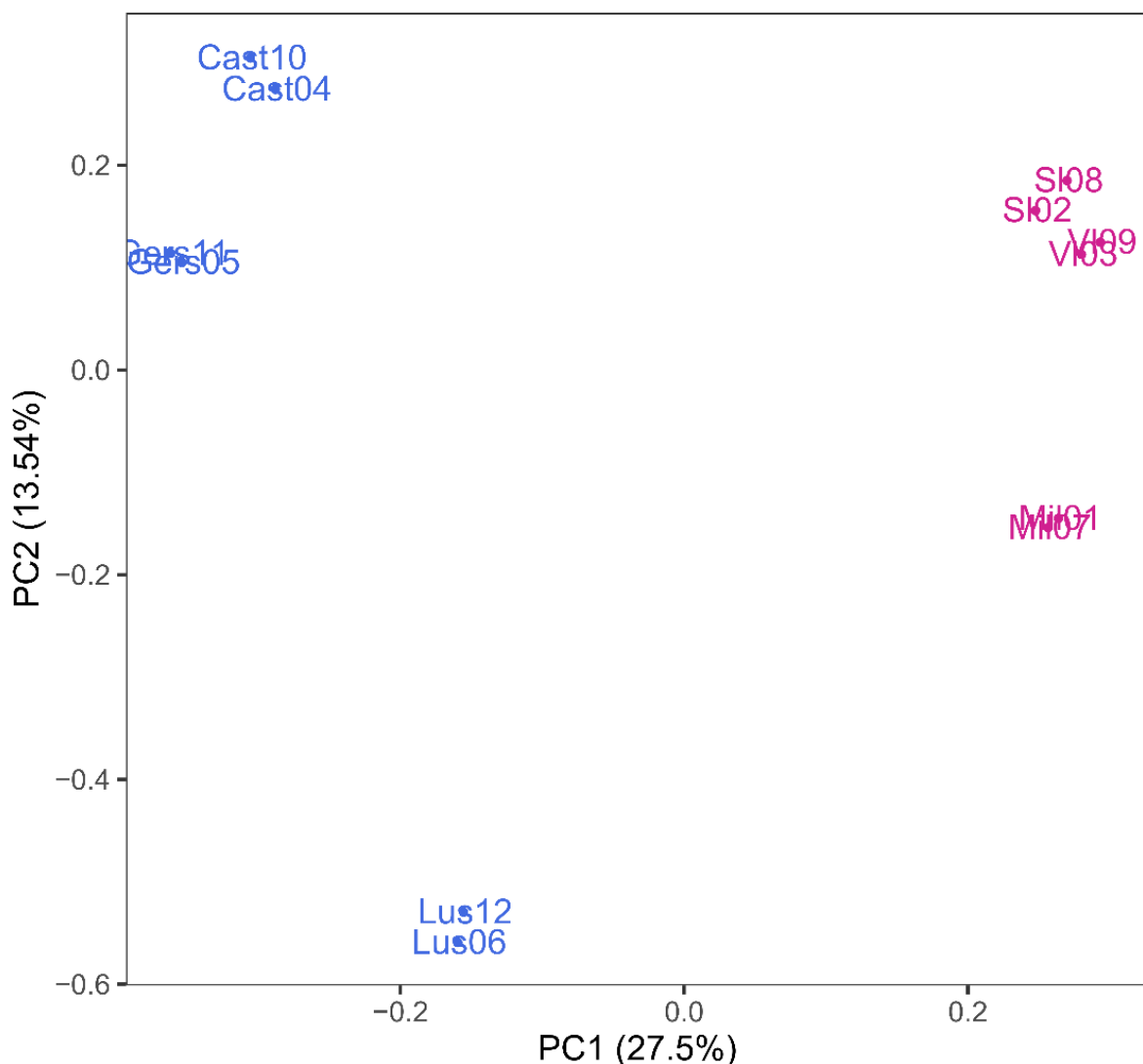
- Chapman NC, et al. 2015. Inheritance of thelytoky in the honey bee *Apis mellifera capensis*. *Heredity* (Edinb) 114:584–592.
- Charif D, Lobry JR. 2007. Seqinr 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. *Structural approaches to sequence evolution: molecules, networks, populations*. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 207–232.
- Christmas MJ, Smith NMA, Oldroyd BP, Webster MT. 2019. Social parasitism in the honeybee (*Apis mellifera*) is not controlled by a single SNP. *Mol Biol Evol.* 36:1764–1767.
- Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* (Austin) 6:80–92.
- Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol.* 23:3133–3157.
- Davis GK. 2012. Cyclical parthenogenesis and viviparity in aphids as evolutionary novelties. *J Exp Zool.* 318B:448–459.
- Dedryver C-A, Bonhomme J, Le Gallic J-F, Simon J-C. 2019. Differences in egg hatching time between cyclical and obligate parthenogenetic lineages of aphids. *Insect Sci.* 26:135–141.
- Dedryver CA, Le Gallic JF, Maheo F, Simon JC, Dedryver F. 2013. The genetics of obligate parthenogenesis in an aphid species and its consequences for the maintenance of alternative reproductive modes. *Heredity* (Edinb) 110:39–45.
- Defardini H, et al. 2023. Evolutionary consequences of loss of sexual reproduction on male-related traits in parthenogenetic lineages of the pea aphid. *Mol Ecol.* 32:3672–3685.
- Fazalova V, Nevado B. 2020. Low spontaneous mutation rate and pleistocene radiation of pea aphids. *Mol Biol Evol.* 37:2045–2051.
- Ferretti L, Ramos-Onsins SE, Perez-Enciso M. 2013. Population genomics from pool sequencing. *Mol Ecol.* 22:5561–5576.
- Frantz A, Plantegenest M, Simon J-C. 2006. Temporal habitat variability and the maintenance of sex in host populations of the pea aphid. *Proc Biol Sci.* 273:2887–2891.
- Frantz A, Plantegenest M, Simon JC. 2009. Host races of the pea aphid *Acyrtosiphon pisum* differ in male wing phenotypes. *Bull Entomol Res.* 100:59–66.
- Gallot A, Shigenobu S, Hashiyama T, Jaubert-Possamai S, Tagu D. 2012. Sexual and asexual oogenesis require the expression of unique and shared sets of genes in the insect *Acyrtosiphon pisum*. *BMC Genomics.* 13:76.
- Garud NR, Messer PW, Petrov DA. 2021. Detection of hard and soft selective sweeps from *Drosophila melanogaster* population genomic data. *PLoS Genet.* 17:e1009373.
- Gautier M, et al. 2018. The genomic basis of color pattern polymorphism in the harlequin ladybird. *Curr Biol.* 28:3296–3302.
- Guyomar C, et al. 2018. Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches. *Microbiome* 6:181.
- Halkett F, Plantegenest M, Bonhomme J, Simon J-C. 2008. Gene flow between sexual and facultatively asexual lineages of an aphid species and the maintenance of reproductive mode variation. *Mol Ecol.* 17:2998–3007.
- Hartfield M, Keightley PD. 2012. Current hypotheses for the evolution of sex and recombination. *Integr Zool.* 7:192–209.
- Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R. 2018. Measuring genetic differentiation from pool-seq data. *Genetics* 210:315–330.
- Hsouna A, Lawal HO, Izevbaye I, Hsu T, O'Donnell JM. 2007. *Drosophila* dopamine synthesis pathway genes regulate tracheal morphogenesis. *Dev Biol.* 308:30–43.
- IAGC. 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8:e1000313.
- Innes DJ. 1997. Sexual reproduction of *Daphnia pulex* in a temporary habitat. *Oecologia* 111:53–60.
- Innes DJ, Dunbrack RL. 1993. Sex allocation variation in *Daphnia pulex*. *J Evol Biol.* 6:559–575.
- Jaquière J, et al. 2014. Genetic control of contagious asexuality in the pea aphid. *PLoS Genet.* 10:e1004838.
- Jaquière J, et al. 2018. Disentangling the causes for faster-X evolution in aphids. *Genome Biol Evol.* 10:507–520.
- Johnson KP, et al. 2018. Phylogenomics and the evolution of hemipteroid insects. *Proc Natl Acad Sci U S A.* 115:12775–12780.
- Joron M, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 477:203–206.
- Keightley PD, et al. 2015. Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Mol Biol Evol.* 32:239–243.
- Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics* 196:313–320.
- Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160:765–777.
- Kofler R, et al. 2011. Popoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One.* 6:e15925–e15925.
- Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.
- Lattorff HMG, Moritz RFA, Crewe RM, Solignac M. 2007. Control of reproductive dominance by the thelytoky gene in honeybees. *Biol Lett.* 3:292–295.
- Lattorff HMG, Moritz RFA, Fuchs S. 2005. A single locus determines thelytokous parthenogenesis of laying honeybee workers (*Apis mellifera capensis*). *Heredity* (Edinb) 94:533–537.
- Le Trionnaire G, et al. 2007. Seasonal photoperiodism regulates the expression of cuticular and signalling protein genes in the pea aphid. *Insect Biochem Mol Biol.* 37:1094–1102.
- Le Trionnaire G, et al. 2012. Transcriptomic profiling of the reproductive mode switch in the pea aphid in response to natural autumnal photoperiod. *J Insect Physiol.* 58:1517–1524.
- Le Trionnaire G, et al. 2019. An integrated protocol for targeted mutagenesis with CRISPR-Cas9 system in the pea aphid. *Insect Biochem Mol Biol.* 110:34–44.
- Le Trionnaire G, Hardie J, Jaubert-Possamai S, Simon JC, Tagu D. 2008. Shifting from clonal to sexual reproduction in aphids: physiological and developmental aspects. *Biol Cell.* 100:441–451.
- Letunic I, Khedkar S, Bork P. 2020. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49:D458–D460.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li B, et al. 2020. A large genomic insertion containing a duplicated follistatin gene is linked to the pea aphid male wing dimorphism. *Elife* 9:e50608.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li YY, Park H, Smith TE, Moran NA. 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Mol Biol Evol.* 36:2143–2156.
- Liu Y, et al. 2016. Whole-animal genome-wide RNAi screen identifies networks regulating male germline stem cells in *Drosophila*. *Nat Commun.* 7:12149.
- Liu H, et al. 2017. Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. *Mol Biol Evol.* 34:119–130.

- Lynch M, Seyfert A, Eads B, Williams E. 2008. Localization of the genetic determinants of meiosis suppression in *Daphnia pulex*. *Genetics* 180:317–327.
- Mark Welch D, Meselson M. 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* 288:1211–1215.
- Martens K, Rossetti G, Horne DJ. 2003. How ancient are ancient asexuals? *Proc Biol Sci*. 270:723–729.
- Mathers TC, et al. 2020. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Mol Biol Evol*. 38:856–875.
- Maynard Smith J. 1971. The origin and maintenance of sex. In: Williams GC, editor. *Group selection*. Chicago: Aldine Atherton. p. 163–175.
- Mérot C, Llaurens V, Normandeau E, Bernatchez L, Wellenreuther M. 2020. Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nat Commun*. 11: 670.
- Moran NA. 1992. The evolution of aphid life cycles. *Annu Rev Entomol*. 37:321–348.
- Muller HJ. 1964. The relation of recombination to mutational advance. *Mut Res*. 1:2–9.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A*. 70:3321–3323.
- Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A*. 76:5269–5273.
- Neiman M, Sharbel TF, Schwander T. 2014. Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. *J Evol Biol*. 27:1346–1359.
- Olazuaga L, et al. 2020. A whole-genome scan for association with invasion success in the fruit fly *Drosophila suzukii* using contrasts of allele frequencies corrected for population structure. *Mol Biol Evol*. 37:2369–2385.
- Oppold AM, Pfenninger M. 2017. Direct estimation of the spontaneous mutation rate by short-term mutation accumulation lines in *Chironomus riparius*. *Evol Lett*. 1:86–92.
- Peccoud J, et al. 2014. Widespread host-dependent hybrid unfitness in the pea aphid complex. *Evolution* 68:2983–2995.
- Peccoud J, Maheo F, De La Huerta M, Laurence C, Simon JC. 2015. Genetic characterisation of new host-specialised biotypes and novel associations with bacterial symbionts in the pea aphid complex. *Insect Conserv Divers*. 8:484–492.
- Peccoud J, Ollivier A, Plantegenest M, Simon JC. 2009. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc Natl Acad Sci U S A*. 106:7495–7500.
- Przeworski M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* 160:1179–1189.
- R Core Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing.
- Rispe C, Pierre J-S. 1998. Coexistence between cyclical parthenogens, obligate parthenogens, and intermediates in a fluctuating environment. *J Theor Biol*. 195:97–110.
- Sabin LR, et al. 2009. *Ars2* regulates both miRNA- and siRNA-dependent silencing and suppresses RNA virus infection in *Drosophila*. *Cell* 138:340–351.
- Sandrock C, Vorburger C. 2011. Single-locus recessive inheritance of asexual reproduction in a parasitoid wasp. *Curr Biol*. 21:433–437.
- Schon I, Martens K, van Dijk P. 2009. Lost sex: the evolutionary biology of parthenogenesis. Dordrecht: Springer.
- Simon J-C, Delmotte F, Rispe C, Crease T. 2003. Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. *Biol J Linn Soc*. 79:151–163.
- Simon JC, Rispe C, Sunnucks P. 2002. Ecology and evolution of sex in aphids. *Trends Ecol Evol*. 17:34–39.
- Simon J-C, Stoeckel S, Tagu D. 2010. Evolutionary and functional insights into reproductive strategies of aphids. *C R Biol*. 333: 488–496.
- Stelzer C-P, Schmidt J, Wiedlroither A, Riss S. 2010. Loss of sexual reproduction and dwarfing in a small metazoan. *PLoS One*. 5: e12854.
- Tucker AE, Ackerman MS, Eads BD, Xu S, Lynch M. 2013. Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. *Proc Natl Acad Sci U S A*. 110:15740–15745.
- van der Kooij CJ, Schwander T. 2014. On the fate of sexual traits under asexuality. *Biol Rev Camb Philos Soc*. 89:805–819.
- Vrijenhoek R, Dawley R, Cole CJ, Bogart J. 1989. A list of known unisexual vertebrates. In: Dawley RM, Bogart JP, editors. *Evolution and ecology of unisexual vertebrates*. New York: New York State Museum. p. 19–23.
- Wallberg A, Pirk CW, Allsopp MH, Webster MT. 2016. Identification of multiple loci associated with social parasitism in honeybees. *PLoS Genet*. 12:e1006097.
- Xu S, et al. 2015. Hybridization and the origin of contagious asexuality in *Daphnia pulex*. *Mol Biol Evol*. 32:3215–3225.
- Yagound B, et al. 2020. A single gene causes thelytokous parthenogenesis, the defining feature of the Cape honeybee *Apis mellifera capensis*. *Curr Biol*. 30:2248–2259.
- Yang S, et al. 2015. Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* 523:463–467.
- Ye Z, Molinier C, Zhao C, Haag CR, Lynch M. 2019. Genetic control of male production in *Daphnia pulex*. *Proc Natl Acad Sci U S A*. 116: 15602–15609.
- Zeng K, Fu YX, Shi S, Wu CI. 2006. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* 174: 1431–1439.

Associate editor: Dr Aida Andres

Supplementary materials for chapter 1

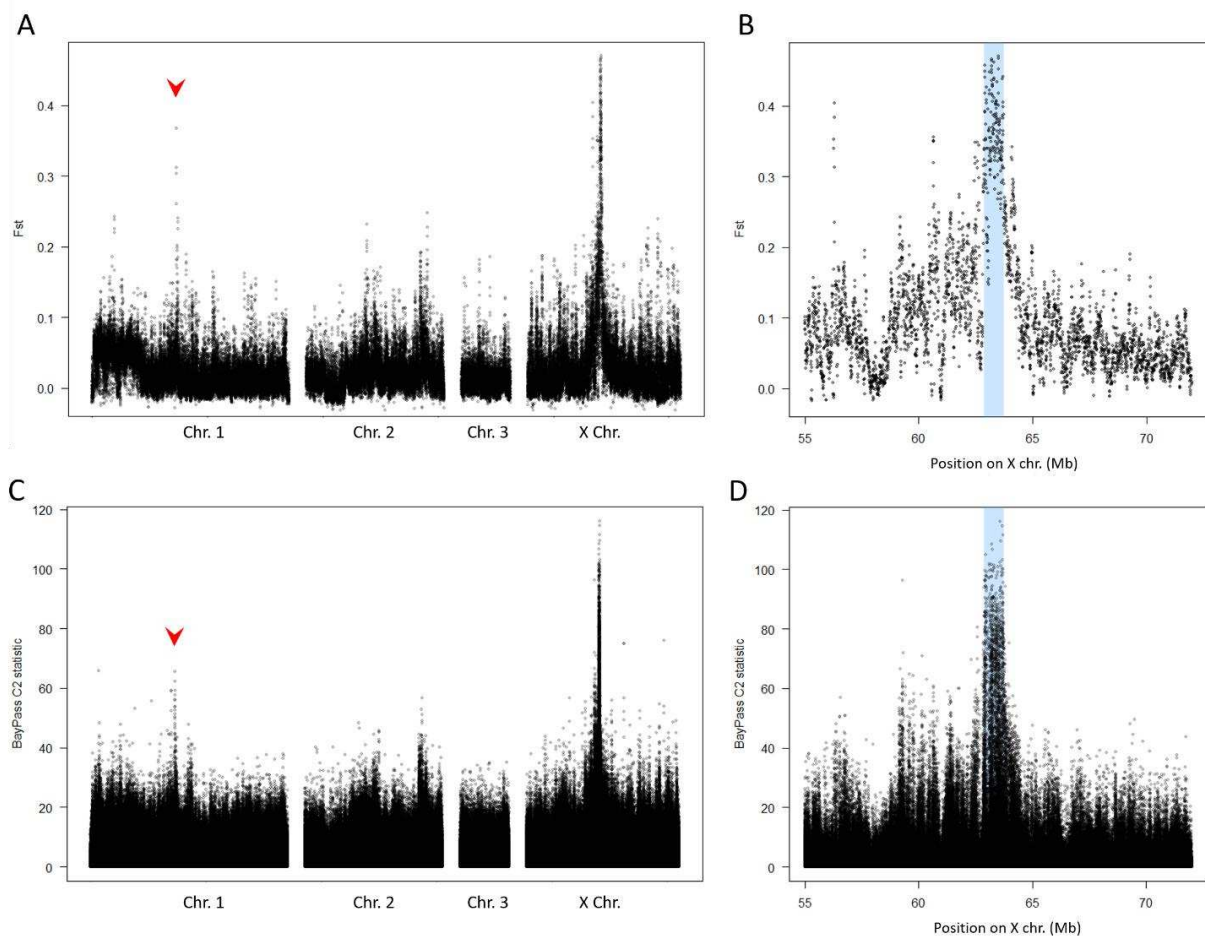
Supplementary File 1. Principal component analysis (PCA) score plot of the first two components calculated on allele frequencies for the 12 libraries (six populations, with two replicates each) constructed from *A. pisum* DNA extracts. Only 50,000 randomly drawn SNPs were used. The results remain unchanged when we repeat this analysis using a set of another 50,000 randomly selected SNPs. OP populations are shown in blue and CP populations in red. The first axis (that accounts for 27.5% of the variance) separates populations by reproductive mode and the second axis discriminates mainly the different OP populations (13.5% of variance explained). Details on the populations are given in Table 1 of the main text.



Supplementary File 2. Pairwise F_{ST} between populations of *A. pisum*. Comparisons between pairs of populations of the same reproductive mode are highlighted in grey. CP: cyclical parthenogenetic populations, OP: obligate parthenogenetic populations. Details on the populations are given in Table 1 of the main text.

	OP_Cast	OP_Gers	OP_Luz	CP_Mil	CP_SI	CP_VI
OP_Cast	-	-0.0447	-0.0197	0.0112	0.0055	0.0104
OP_Gers	-	-	-0.0261	0.0138	0.0152	0.0156
OP_Luz	-	-	-	-0.0006	0.0052	0.0059
CP_Mil	-	-	-	-	-0.0106	-0.0118
CP_SI	-	-	-	-	-	-0.0137
CP_VI	-	-	-	-	-	-

Supplementary File 3. Detection of genomic regions associated with reproductive mode variation along the genome of *Acyrtosiphon pisum*. A) Genetic differentiation (F_{ST}) between obligate parthenogenetic (OP) and cyclical parthenogenetic (CP) populations (20-kb windows sliding by 5-kb steps). B) Detail of the portion of the X chromosome that contains the main outlier region. C) $C2$ statistic from BayPass for each SNP for the whole genome and (D) for the main outlier region. The $C2$ statistic contrasts allele frequencies between OP and CP populations while correcting for population structure, to identify SNPs associated with this binary trait. The 840-kb region identified as the main candidate is shown in blue. The red arrow identifies a 30 kb region misplaced in the v3.0 genome assembly (Li, et al. 2019), which is actually located at 2 kb from the main outlier region in the middle of the X chromosome. This misassembled region (see also Supplementary Files 4 and 5 for more details) was moved to its expected position in Figure 1.



Supplementary File 4. This supplementary file describes the new data that were sequenced to build new genome assemblies for the OP and CP lineages of the pea aphid, the construction of these assemblies and the analyses that were performed on the 840-kb candidate region with these genomes.

OP genome sequencing and assembly

We sequenced and assembled an OP genome with long reads (Oxford Nanopore, ONT). The lineage selected for sequencing (lineage X6-2 from cross 6 in the Supplementary Figure S1 from Jaquiéry, et al. 2014) was characterized as OP in conditions inducing sexual morph production. DNA was extracted from parthenogenetic embryos dissected from 20 adult females of this X6-2 OP lineage. Embryos were ground in lysis buffer using a potter. The lysis solution was then used in a Phenol protocol using high vacuum grease silicone to separate aqueous and organic phases without pipetting, and low centrifuge speed to preserve DNA integrity. The DNA was directly precipitated with isopropanol and then resuspended in ultra-pure sterilized water. DNA concentration was measured with Qubit fluorometer and DNA fragment length (range 17kb-100kb) was assessed with both a Pulse Field Gel Electrophoresis and a Femto Pulse Run.

A Nanopore genomic DNA library was then prepared using the Ligation Sequencing Kit (Oxford Nanopore Technologies), following the manufacturers protocol, and sequenced on an R9.4 flow cell for 72 hours (ONT GridION technology). See suppl. Table 1 (in the present file) for a summary of the characteristics of sequence data obtained, which are publicly available on NCBI (BioProject ID PRJNA745262).

The Nanopore reads were first trimmed with PoreChop v0.2 (Wick, et al. 2017), using default parameters. Reads were then assembled with wtdbg v2.5 (Ruan and Li 2020) with the options -x ont -g 500m and polished with the same tools after an alignment step of the trimmed reads with minimap v2.14 (Li 2018). The characteristics of this OP genome assembly are summarized in Table 2 (this file). This genome assembly is publicly available at https://bipaa.genouest.org/sp/acyrthosiphon_pisum/download/genome/OP/.

Improved CP genome sequencing and assembly

As the genome available for a CP lineage of the pea aphid (reference v3.0, Li, et al. 2019) contains small assembly errors – which could have affected our conclusions if the candidate region was involved – we chose to build a new genome assembly for a CP lineage. To do so, we combined PacBio and ONT sequencing data with an optical map performed on a CP lineage of the pea aphid (clone LSR1).

For the optical map, ultra-high molecular weight (uHMW) DNA was purified from 0.2 g of frozen larvae from the LSR1 aphid lineage according to the Bionano Animal Tissue DNA Isolation Grinding Protocol (800002 - Bionano Genomics) with the following specifications and modifications. Briefly, the aphid larvae were disrupted in the homogenization buffer with a potter. Nuclei were washed and then embedded in agarose plugs. After overnight proteinase K digestion in the presence of Lysis Buffer (Bionano Genomics) and one hour treatment with RNase A (Qiagen), plugs were washed four times in 1x Wash Buffer (Bionano Genomics) and five times in 1x TE Buffer (ThermoFisher Scientific). Then, plugs were melted two minutes at 70°C and solubilized with 2 µL of 0.5 U/µL AGARase enzyme (ThermoFisher Scientific) for 45 minutes at 43°C. A dialysis step was performed in 1x TE Buffer

(ThermoFisher Scientific) for 45 minutes to purify DNA from any residues. The DNA samples were quantified using the Qubit dsDNA BR Assay (Invitrogen). Megabase-sized DNA fragments were visualized by pulsed field gel electrophoresis (PFGE).

Labeling and staining of the uHMW DNA were performed according to the Bionano Prep Direct Label and Stain (DLS) protocol (30206 - Bionano Genomics). Briefly, labeling was performed by incubating 750 ng genomic DNA with 1× DLE-1 Enzyme (Bionano Genomics) for two hours in the presence of 1× DL-Green (Bionano Genomics) and 1× DLE-1 Buffer (Bionano Genomics). Following proteinase K digestion and DL-Green cleanup, the DNA backbone was stained by mixing the labeled DNA with DNA Stain solution (Bionano Genomics) in presence of 1× Flow Buffer (Bionano Genomics) and 1× DTT (Bionano Genomics), and incubated overnight at room temperature. The DLS DNA concentration was measured with the Qubit dsDNA HS Assay (Invitrogen).

Labelled and stained DNA was loaded on the Saphyr chip. Loading of the chip and running of the Bionano Genomics Saphyr System were all performed according to the Saphyr System User Guide (30247 - Bionano Genomics). Data processing was performed using the Bionano Genomics Access software (<https://bionanogenomics.com/support-page/bionano-access-software/>). A total of 1.1 Tb of data was generated. From this data, molecules with a size larger than 150 kb, the threshold for map assembly, represent 356 Gb of data. These filtered data (> 150 kb), corresponding to 651x coverage of the 550 Mb estimated size of *A. pisum* genome, were compiled from 1,557,530 molecules with N50 of 223 kb and an average label density of 13.2/100kb. The filtered molecules were aligned using RefAligner with default parameters. It produced 127 genome maps with a N50 of 20 Mb for a total genome map length of 778.5 Mb. As the map size was longer than expected, due to the heterozygosity, we purged it in order to obtain only one haplotype for each optical map before the hybrid scaffolding step. For that, we first used runCharacterise from Bionano tools to align maps with each other, and created an alignment file (xmap file). From that file, we recovered supernumerary maps which align globally to other maps, with an in-house java program. This way, we purged the 778.5 Mb optical maps and obtained a genome map length of 527 Mb (consistent with the pea aphid genome size), consisting in 34 maps.

For PacBio sequencing, high molecular weight DNA was extracted from the CP lineage (LSR1) following a protocol similar to the one used for ONT sequencing of the OP clone (see above). PacBio genomic DNA libraries were prepared using the SMRTbell Template Prep Kit 1.0 (PacificBiosciences) following the manufacturer's protocol and sequenced with six SMRTCells 1M on PacBio Sequel (PacificBiosciences) in Gentyane Platform (Clermont-Ferrand, France) and Centre for Genomic Research (University of Liverpool, UK). The raw data (see Table 1 below) are publicly available on NCBI (BioProject ID PRJNA745262). PacBio raw reads were first treated to produce CCS with the ccs program of the suite PacificBioSystems Pitchfork v3.0 (commit 96f0b06, <https://github.com/PacificBiosciences/pitchfork>) with the option --maxLength 20000. The resulting CCS and subreads from reads with no CCS were mixed. ONT sequencing was also performed on the LSR1 CP lineage, using the same protocol as described above for the OP lineage.

ONT reads and PacBio subreads were then aligned to the *Buchnera aphidicola* strain APS complete genome (NC_002528.1) with minimap2 v2.17 (Li 2018) with default parameters. The reads matching *B. aphidicola* genome were removed for further analyses. The final set of reads was then assembled and polished with flye v2.7.1 (Kolmogorov, et al. 2019) following the website instructions (<https://github.com/fenderglass/Flye>) for a mix of ONT and PacBio sequences (assembly of the 2 sets

with the options -g 530m - iterations 0, then polishing with PacBio data only with the options -resume from polishing - genome-size 530m).

Then, the 34 reduced optical maps were compared individually to the genome sequences with the hybridScaffold.pl script from Solve3.6.1_11162020 (with the options -B 2 and -N 2), and finally all the sequences resulting of hybrid maps were merged into one genome sequence. Lastly this genome sequence was gapfilled with LR_Gapcloser v1.1 (Xu, et al. 2018) (with the option -s p) and the reads corrected with CANU v1.9 (Koren, et al. 2017).

We refer to this final assembly as the “improved CP assembly” (see suppl. Table 2 below for its main characteristics), which is available at https://bipaa.genouest.org/sp/acyrthosiphon_pisum/download/genome/LSR1_CP/.

Supplementary Table 1. Sequencing data used for genome assemblies

	PacBio data	ONT (Nanopore)	
Aphid lineage	LSR1 (cyclically parthenogenetic)	LSR1 (cyclically parthenogenetic)	X6_2 (obligately parthenogenetic)
Number of reads	8,007,204	372,636	1,330,350
Sequenced bases	51,786,985,914	4,240,655,565	15,864,369,067
Genome coverage	97.7X	8X	30X
Mean read length (bp)	6,468	11,380	11,924
Median read length (bp)	6,163	9,010	10,262
N50 (bp)	8,750	17,775	16,988

Supplementary Table 2. Genome assembly statistics. Busco analyses were realized with BUSCO 4.0.6, using the dataset insecta_odb10 (1367 BUSCOs, https://busco.ezlab.org/list_of_lineages.html).

	Improved CP genome	OP genome
Aphid lineage	LSR1 (cyclically parthenogenetic)	X6_2 (obligately parthenogenetic)
Data used	PacBio, Nanopore, Optical map	Nanopore
Number of scaffolds	32	3889
Assembly size (Mbp)	527	478
N50 (Mbp)	60	0.955
L50	4	127
BUSCO - single	1220	902
BUSCO - Duplicated	63	8
BUSCO - fragmented	13	186
BUSCO - Missing	71	271

Identification of the actual chromosomal localization of the 30-kb outlier region found on chromosome 1 on the v3.0 reference genome

Our genome scan identified two main genomic regions with high F_{ST} values, the main 840-kb X-linked candidate region and a short one (30 kb) on chromosome 1 (located between 73,118,603 and 73,151,851, see Supplementary File 3). Given the abrupt changes in F_{ST} at the border of this 30-kb region, we suspected a genome assembly error. In a previous study, the entire *A. pisum* genome was assigned to the X or autosomes based on ratios of sequencing depth in males (X0) to females (XX) (Jaquiéry, et al. 2018). Using these data, we discovered that the 30-kb region was indeed misplaced on the v3.0 reference genome (Li, et al. 2019) and actually belongs to the X chromosome. We located the 30-kb chr1 region on Super-Scaffold_100003 of the improved CP genome, between positions 20,170,923 and 20,205,127 (see Supplementary File 5A). Super-Scaffold_100003 corresponds to a region on the X chromosome of the v3.0 reference genome (Li, et al. 2019) and the 30-kb region would be localized between positions 60,702,513 and 60,707,859. This places the short region of high F_{ST} at ~2 Mb of the main 840-kb candidate region (which also locates on Super-Scaffold_100003, see Supplementary File 5A). These regions may thus be under the influence of the same locus controlling reproductive mode. However, since the F_{ST} values of this 30-kb region were below the F_{ST} threshold of 0.4, we did not further retain it as a candidate region.

Structure of the 840-kb candidate region in OP and CP genomes

Finally, to characterize the genomic structure of the 840-kb candidate region, we located this region in the OP genome and in the improved CP genome. In the improved CP genome, we located the 840-kb candidate region on the Super-Scaffold_100003, between positions 22,415,764 and 23,247,663 (Supplementary File 5A). In the OP genome, the whole 840-kb candidate region was also found on a single scaffold (between positions 249,000 and 1,093,000 on scaffold cgt7). Pairwise alignments of the sequence corresponding to the 840-kb region (flanked by 200 kb on each side) in the 3 different genome assemblies (the v3.0 reference genome, the improved CP genome and the OP genome) were assessed using NUCmer v3.07 from package MUMmer v3.22 (Kurtz, et al. 2004). Alignments were filtered using the script delta-filter to keep optimal correspondence with a minimum length of 1000 bp and a minimum alignment identity of 90. We found that the local assembly of the region corresponding to the 840-kb region was concordant between the improved CP assembly and the v3.0 reference assembly and also with the OP genome assembly (Supplementary File 5B and 4C, see also Figure 4 in the manuscript).

References

- Jaquiéry J, Peccoud J, Ouisse T, Legeai F, Prunier-Leterme N, Gouin A, Nouhaud P, Brisson JA, Bickel R, Purandare S, et al. 2018. Disentangling the causes for faster-X evolution in aphids. **Genome Biology and Evolution** 10:507-520.
- Jaquiéry J, Stoeckel S, Larose C, Nouhaud P, Risper C, Mieuzet L, Bonhomme J, Maheo F, Legeai F, Gauthier JP, et al. 2014. Genetic control of contagious asexuality in the pea aphid. **Plos Genetics** 10.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. **Nature Biotechnology** 37:540-546.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. **Genome Research** 27:722-736.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. **Genome Biology** 5:R12.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. **Bioinformatics** 34:3094-3100.

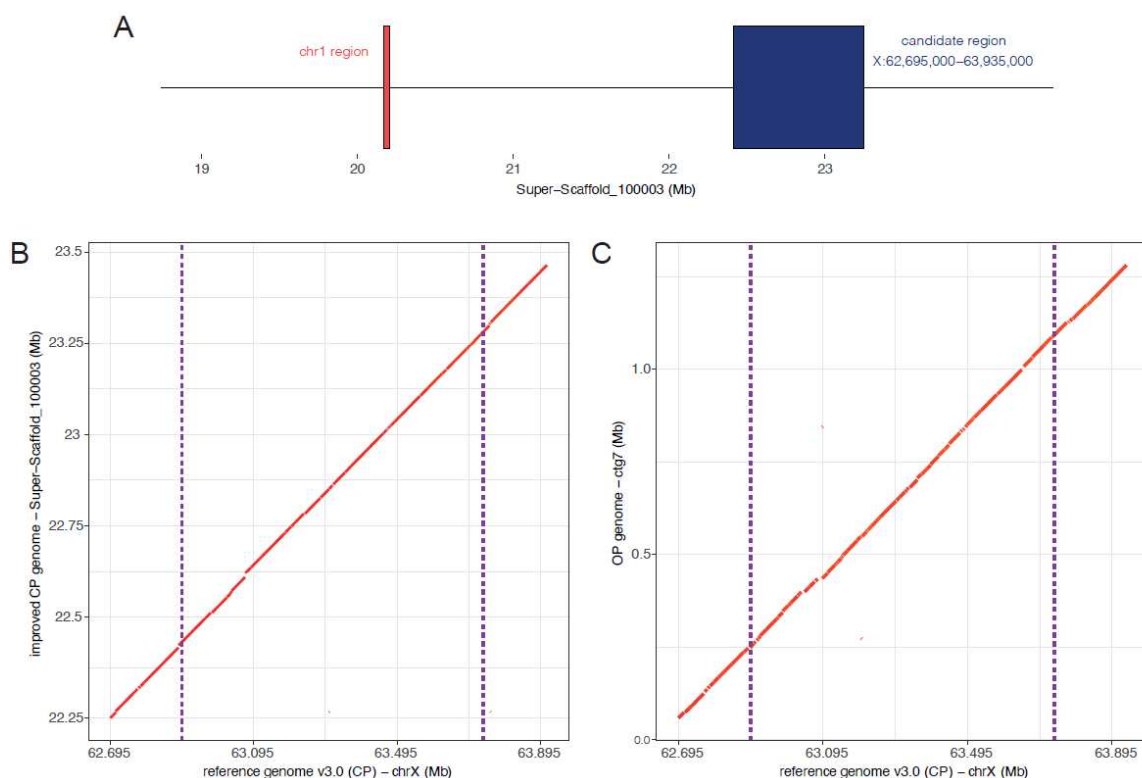
Li YY, Park H, Smith TE, Moran NA. 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. **Molecular Biology and Evolution** 36:2143-2156.

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. **Nature Methods** 17:155-158.

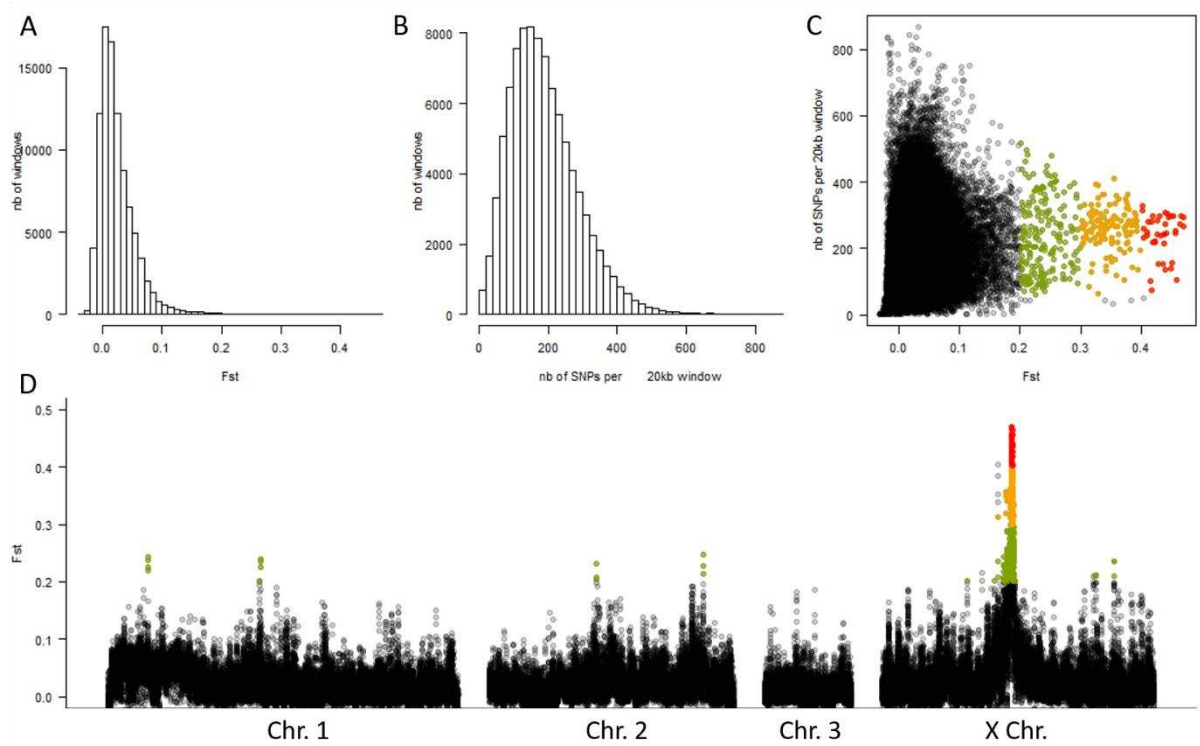
Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. **PLoS Computational Biology** 13:e1005595.

Xu G-C, Xu T-J, Zhu R, Zhang Y, Li S-Q, Wang H-W, Li J-T. 2018. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. **GigaScience** 8.

Supplementary File 5. A) Location of the 30-kb chromosome 1 region and of the 840-kb candidate region on Super-Scaffold_100003 from the improved CP genome assembly. B and C) MUMmer alignment plots for the nucleotide sequences from the different genome assemblies corresponding to the 840-kb candidate region (plus 200 kb on each side). The x-axis represents coordinates of the v3.0 reference genome assembly. The y-axis represents coordinates of Super-Scaffold_100003 from the improved CP genome assembly (B) or of scaffold ctg7 from the OP genome assembly (C). The purple vertical dashed lines delimit the 840-kb candidate region.



Supplementary File 6. Distribution of F_{ST} (A) and number of SNPs (B) per 20-kb window sliding by 5-kb steps, and their relationships (C). In panels C and D, only windows with at least 50 SNPs and with F_{ST} above a given threshold were coloured (green for $F_{ST} > 0.2$, yellow for $F_{ST} > 0.3$ and red for $F_{ST} > 0.4$). The secondary peak just to the left of the main 840-kb candidate region (comprising four grey dots in panel D) is composed of windows with a low number of SNPs (see panel C, bottom right), which were also not supported by the BayPass $C2$ statistic.

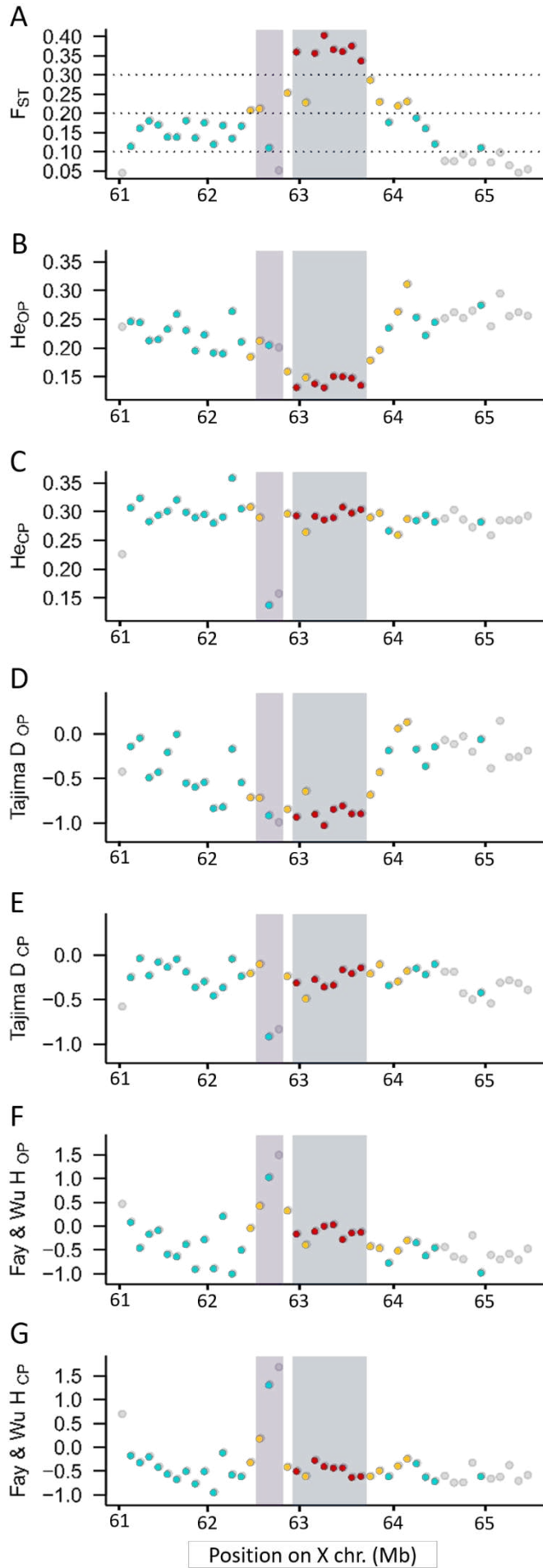


Supplementary File 7. Population genetic indices calculated along a region of the X chromosome of *Acyrtosiphon pisum* in non-overlapping 100 kb windows. This region, ranging from 61 to 65.5 Mb, encompasses the main candidate region associated with reproductive mode variation (highlighted in blue) and a locus called *aphicarus* associated with wing polymorphism in males in violet (Braendle, et al. 2005; Li, et al. 2020). A) F_{ST} between OP and CP populations; B and C) heterozygosity in OP and CP populations, respectively; D and E) Tajima's D for OP and CP populations, respectively; F and G) Fay & Wu's H for OP and CP populations, respectively. Each point (a 100 kb window) was colored according to its F_{ST} value (red for windows with an average F_{ST} above 0.3, yellow for F_{ST} between 0.2 and 0.3, blue for those with F_{ST} between 0.1 and 0.2, grey for those below 0.1 to facilitate the visualization of the possible relationship between regions with high F_{ST} values and their values at other indices. *Aphicarus* is characterized by low differentiation (F_{ST}) between OP and CP populations, and similar patterns for both population types at other indices, namely reduced heterozygosity and Tajima's D and increased Fay and Wu's H. On the other hand, particular signatures were only detected in the OP populations for the candidate region associated with the variation in reproductive mode.

References

Braendle C, Caillaud MC, Stern DL. 2005. Genetic mapping of *aphicarus* – a sex-linked locus controlling a wing polymorphism in the pea aphid (*Acyrtosiphon pisum*). **Heredity** 94:435-442.

Li B, Bickel RD, Parker BJ, Saleh Ziabari O, Liu F, Vellichirammal NN, Simon J-C, Stern DL, Brisson JA. 2020. A large genomic insertion containing a duplicated follistatin gene is linked to the pea aphid male wing dimorphism. **Elife** 9:e50608.



Supplementary File 8. Description and annotation of the 32 genes and of the variants with moderate to high impact in the 840-kb candidate region.

See *SupplementaryFile8.xlsx* on

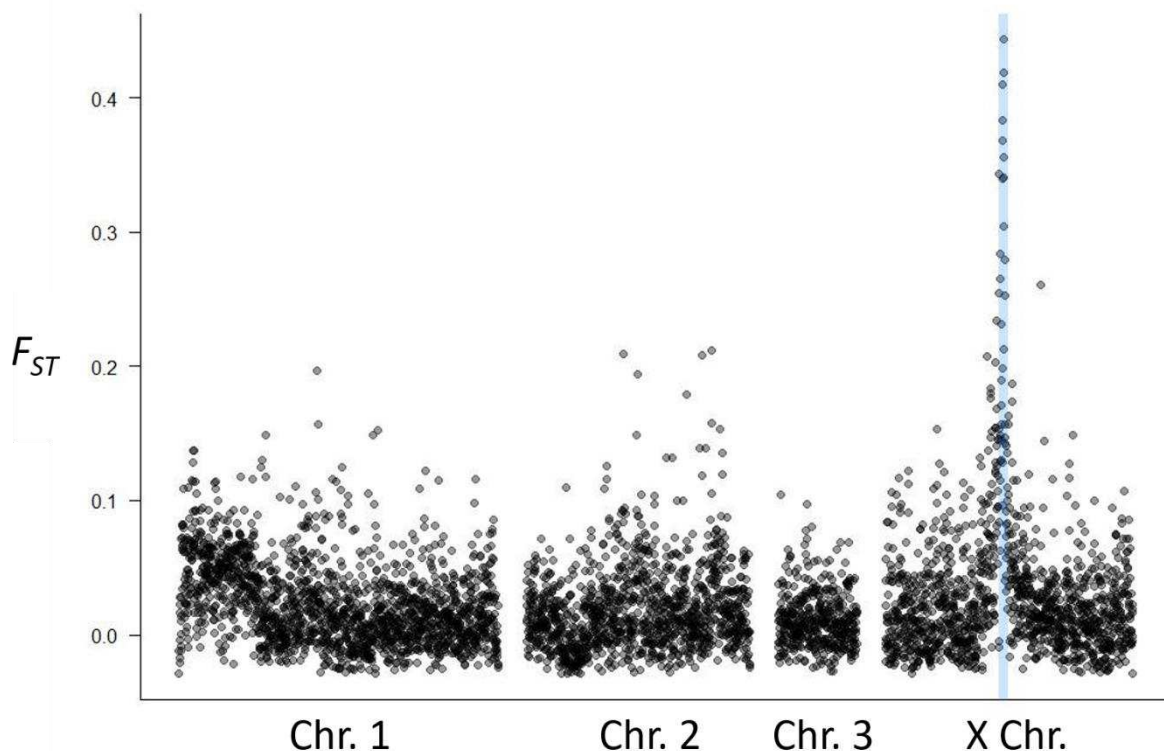
<https://academic.oup.com/gbe/article/15/9/evad168/7275732#supplementary-data>

Supplementary File 9. Summary of the coverage of the 32 genes located in the candidate region in OP and CP populations.

See *SupplementaryFile9.xlsx* on

<https://academic.oup.com/gbe/article/15/9/evad168/7275732#supplementary-data>

Supplementary File 10. Detection of genomic regions associated with reproductive mode variation in *Acyrtosiphon pisum* based on the genetic differentiation (F_{ST}) between obligate parthenogenetic (OP) and cyclical parthenogenetic (CP) populations (100-kb non-overlapping windows). Here, we restricted our analysis to the 181,204 tri-allelic SNPs for which the third allele was supported by only one read. For F_{ST} calculation, this third allele was ignored and the SNPs were considered biallelic. The 840-kb region identified as candidate using the full dataset is shown in blue. This subset of data composed of tri-allelic SNPs therefore gave results perfectly similar to those obtained with the whole dataset, showing that the inclusion of these 181,204 SNPs does not affect our conclusions.



Chapter 2

After having characterized the candidate region for sex loss in one race of the pea aphid, I set out to identify the genetic basis of sex loss along a continuum of divergence from two other closely related races of the pea aphid to a more divergent species, the green peach aphid, in order to get insight into the evolutionary history of sex loss in this group. This chapter is a first draft of a paper that requires more work before submission.

To each his own: no evidence of shared genetic basis for sex loss among different aphid species

Hélène Defendini¹, Chris Bass², Benoit Barrès³, Lise Roy⁴, Maud Rimbault¹, Lucie Mieuze¹, Romuald Cloteau¹, Gaëtan Denis¹, Frédérique Mahéo¹, Natalie Prunier-Leterme¹, Stéphanie Robin^{1,3}, Fabrice Legeai^{1,5}, Jean-Christophe Simon¹, Julie Jaquiéry¹

¹ UMR 1349 IGEPP, INRAE, Institut Agro, Université de Rennes, 35653 Le Rheu, France

² College for Life and Environmental Sciences, University of Exeter, Penryn, Cornwall, UK

³ ANSES/INRAE - USC CASPER Lyon, France

⁴ CEFE, University of Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France

⁵ UMR 1349 IGEPP, INRAE, Institut Agro, Université de Rennes, 35000 Rennes, France

Abstract

Evolutionary transitions to asexual reproduction are common in eukaryotes, but the genetic basis of sex loss is poorly understood. Aphids typically reproduce by cyclical parthenogenesis (CP), alternating several successive clonal generations with one sexual generation. However, CP lineages frequently co-exist within the same aphid species with derived lineages that undergo obligate parthenogenesis (OP). In a previous work we showed that a single 840kb locus containing 32 genes controls sex loss in the alfalfa-adapted race of the pea aphid complex, *Acyrtosiphon pisum*. In this study, we sought to identify the genetic basis of sex loss in different aphid taxa located on a divergence gradient by comparing the genomes of their respective CP and OP lineages. In particular, we analyzed genomic differences between CP and OP lineages in two other host races of the pea aphid, specialized respectively on pea and clover, as well as in the distant *Myzus persicae* species. We were able to detect peaks of differentiation associated with sex loss in the different taxa, but they were located in distinct genomic regions and none of them showed homology with the previously characterized locus in the alfalfa-adapted race of *A. pisum*. This result suggests that sex loss in aphids can be achieved by several mechanisms, consistent with the many steps involved in the change between clonality and sexuality in CP lineages. However, some of the observed differentiation peaks may also be linked to genetic differences between CP and OP lineages that are not directly related to sex loss, but rather to its side effects.

Keywords: cyclical parthenogenesis, obligate parthenogenesis, asexuality, reproductive mode transition, GWAS, poolseq, genome resequencing.

INTRODUCTION

While sexuality is the primary mode of reproduction in metazoans, parthenogenesis - where an embryo develops from an unfertilised egg - occurs in various branches of the animal kingdom (Bell, 1982). This phenomenon is observed in a diverse range of groups, including molluscs, insects, crustaceans, nematodes, fish, and reptiles (Schön et al., 2009; Vrijenhoek et al., 1989). The prevalence of sex in evolution is attributed to its long-term advantages, which include purging deleterious mutations and combining advantageous ones in the same genome (Barton & Charlesworth, 1998; Kondrashov, 1993; Muller, 1964). However, it is unclear how sexual lineages can persist in the short term in the face of newly emerging asexual lineages. This is because, under the “all else being equal” hypothesis, sexual organisms are expected to experience a two-fold cost of sex (Lehtonen et al., 2012; Maynard Smith, 1971). An improved understanding of how a new asexual lineage can emerge could help resolve this paradox.

In animals, the loss of sexual reproduction can occur through various mechanisms, such as interspecific hybridization, microbial infection, spontaneous mutations, or the dissemination of contagious asexual factors (Simon et al., 2003). However, the genomic regions underlying the loss of sex are poorly characterized because recombination-based approaches are only possible in the limited number of species with partial loss of sex. The few existing studies have shown that the genetic architecture for these transitions can be simple, consisting of as few as one locus (e.g. Stelzer et al., 2010 in a rotifer, Sandrock & Vorburger, 2011 and Yagound et al., 2020 in hymenopterans, Jaquiéry et al., 2014 in an aphid) or several loci (Sperling et al., 2023 in a *Drosophila*, Tucker et al., 2013 and Xu et al., 2015 in *Daphnia*, Dedryver et al., 2013 in an aphid). However, well-supported candidate genes have only been identified in two species, the Cape honeybee and *Drosophila mercatorum*. In the Cape honeybee, GB45239 has been identified as a candidate gene that putatively contributes to chromosome segregation. However, it is not yet clear whether this gene acts through its protein sequence, its level of expression, or a combination of both (Yagound et al., 2020). Sperling et al. (2023) discovered that a combination of three genes in the facultative parthenogenetic species *D. mercatorum* could induce parthenogenesis in the non-parthenogenetic species *D. melanogaster*, through manipulation of their expression. These genes are *Myc*, which broadly influences the expression of genes involved in cell growth, size, metabolism, apoptosis, and autophagy; *Desat2*, a desaturase; and a mitotic protein kinase known as *polo*. Collectively, these genes play essential roles in the parthenogenetic development of *D. mercatorum*, from oocyte to embryonic development and tissue proliferation.

The limited number of identified genes responsible for sex loss and the likely divergence in the underlying mechanisms among distant taxa render testing for convergence in genes or mechanisms irrelevant. Nonetheless, some groups exhibit a notably elevated incidence of taxa displaying partial or complete loss of sexual reproduction, rendering them prime candidates for investigating the evolutionary trajectory of sex loss. They could be particularly relevant for testing whether similar mechanisms or genes are involved, and the extent to which the evolution towards obligate asexuality is constrained. The aphid group possesses such characteristics, with approximately 40% of the 5,000 described aphid species exhibiting partial or complete loss of sexual reproduction (Moran, 1992). The ancestral reproductive mode employed by aphids is cyclical parthenogenesis (CP), a distinctive life cycle that emerged in the early stages of aphid evolution, around 250 million years ago (Dixon, 1998; Hales et al., 1997). Cyclical parthenogenesis involves alternating periods of apomictic parthenogenesis

during spring and summer (by viviparous parthenogenetic females), followed by a sexual phase involving oviparous sexual females and males during autumn. It is the reduction in daylength typical of the fall that triggers the development of sexual forms (Le Trionnaire et al., 2008), which are also produced through apomictic parthenogenesis, except that one X chromosome is discarded to form XO males (the two female morphs are XX). The sexually produced overwintering eggs, which serve as the sole frost-resistant stage in the aphid's annual life cycle (Simon et al., 2002), produce viviparous parthenogenetic females in the subsequent spring. Interestingly, some aphid lineages have completely lost the sexual phase of the life cycle, resulting in obligate parthenogenesis (OP) (Simon et al., 2010). Many aphid species display polymorphism in their life cycle, which includes CP lineages as well as lineages (referred to as OP) that cannot generate sexual females and hence cannot accomplish the full cycle comprising the two reproductive phases. OP and CP lineages tend to occupy varying habitats (e.g. Defendini et al., 2023; Risper et al., 1998). Only CP lineages can survive in cold regions during the winter because they sexually produce diapausing eggs. In contrast, OP lineages have an advantage in regions with mild winters due to their ability to continue developing throughout the year.

While the transition from clonal to sexual reproduction in CP aphids is triggered by changes in photoperiod, the lack of production of sexual forms in OP aphids is genetically determined, with various lineages exhibiting different degrees of investment in sexual reproduction when raised under sex-inducing conditions (Blackman, 1971; Dedryver et al., 1998, 2001; Delmotte et al., 2002; Halkett et al., 2008; Helden & Dixon, 2002; Simon et al., 1991, 1999; Vorburger et al., 2003). There is no evidence to suggest that microorganisms are the cause of sex loss in aphids (Simon et al., 2007, 2011). A study involving a small number of F1 and F2 offspring obtained from crosses between OP and CP lineages of the peach-potato aphid *Myzus persicae* showed that CP was dominant and OP recessive, and indicated that an epistatic modifier was required to explain the inheritance patterns of the observed phenotypes (Blackman, 1972). Similar investigations conducted on the cereal aphid *Sitobion avenae* revealed that at least two autosomal loci are required to account for the observed F1 and F2 phenotypes (Dedryver et al., 2013). In the pea aphid *Acyrtosiphon pisum*, analyses indicate that sex loss is associated with a single locus on the X chromosome and that the OP allele is recessive (Jaquiéry et al., 2014). Genome scan analysis of multiple microsatellite loci, followed by whole-genome sequencing in pools of over 100 OP and CP lineages, pinpointed the same region, which can be refined to an 840 kb candidate region comprising 32 predicted genes (Jaquiéry et al., 2014; Rimbault et al., 2023). Then, the combination of transcriptomic and protein sequence analyses enabled the identification of eight candidate genes from the initial 32 (Rimbault et al., 2023, Hugué et al., *submitted*). These genes may impact sexual morph production through differential expression in response to photoperiodic signals, altered protein sequences in OP lineages or a combination of both mechanisms. However, the putative functions (when available) of these eight genes provide limited insights into which one is responsible for the loss of sexual reproduction. Given the intricate molecular cascade underpinning the induction of the sexual phase in aphids, there are many potential ways to lose sex, spanning from an absence of photoperiodic signal perception to impaired signal transduction and a failure to respond to the inducing signal during embryonic development.

Here, our objective was to explore the genetic architecture of sex loss - namely the number of loci and their genomic distribution - in various aphid species displaying reproductive polymorphism. Additionally, we aimed to distinguish various evolutionary scenarios that may explain the loss of sex in this group. Firstly, it is possible that the loss of sex occurred long time ago and the reproductive polymorphism was inherited through speciation events, resulting in a trans-specific polymorphism such as the ABO blood group in primates (Ségurel et al., 2012). In this scenario, the same genomic

region would be implicated in sex loss across different species, and the phylogenetic tree of this candidate region would group alleles based on their mode of reproduction rather than by species identity. Secondly, hybridization events may have transferred an asexual allele from one species to another, resulting in an intermediate placement of the candidate region on the phylogeny. Thirdly, particular regions of the genome could be more prone to producing a new allele that confers asexuality due to their genic content. In such instances, the genomic region may be shared among species, but the tree would sort alleles solely by the species. Finally, it is possible that each species lost sex independently, resulting in the identification of different regions according to species. Any recurrence in gene function would be of great interest.

Our prior research unequivocally pinpointed a single genomic region associated with sex loss in the alfalfa host race of the pea aphid (Jaquiéry et al., 2014; Rimbault et al., 2023). This species complex, which includes 15 host races each adapted to distinct legume species (Peccoud et al., 2009), is ideal for addressing our queries, as several host races also display reproductive polymorphism (Frantz et al., 2006). We performed a comprehensive sampling of OP and CP lineages in two additional host races of the pea aphid. We sequenced CP and OP genomes in pools to identify candidate regions for sex loss in each of these host races. Subsequently, we further expanded the range of our analysis by examining the genetic basis of sex loss in a more remote species (*Myzus persicae*) using individual and pool genome resequencing of CP and OP lineages. These new data, coupled with a re-examination of the *A. pisum* alfalfa host race data, suggest multiple genetic causes for sex loss, shedding light on why obligate asexuality is pervasive among aphids.

MATERIALS ET METHODS

Sample collection

We first reanalyzed the *A. pisum* dataset of the alfalfa race that was previously used to conduct genome scans based on pool-sequencing (Jaquiéry et al., 2014; Rimbault et al., 2023). Briefly, these data consist of 3 OP populations collected in mild-winter regions (south-west France), and 3 CP populations collected in cold-winter regions (north-east France, Table 1).

To investigate the genetic basis of sex loss in other host races of the pea aphid complex, aphids were collected from two other host plants, red clover (*Trifolium pratense*) and pea (*Pisum sativum*) in north-eastern and north-western France. As already described in Defendini et al. (2023), only 219 out of the 563 parthenogenetic females from the clover race successfully established a colony in the lab because more than 50% of wild caught individuals were parasitized by entomopathogenic fungi or parasitoid wasps. These 219 lineages were then genotyped at 15 polymorphic microsatellite loci (AIA09M, AIB07M, AIB08M, AIB12M, ApF08M, ApH08M, ApH10M, X111775, X114782, X117294, X118775, X118777, X126104, X127526, X128001, Caillaud et al., 2004; Jaquiéry et al., 2012; Nouhaud et al., 2014) following DNA extraction and PCR protocols described in Peccoud et al. (2008). Lineages that had the same genotype at all 15 loci or differed by at most one allele were identified with the R package RClone (Bailleul et al., 2016) and were assumed to be clonal copies of the same multilocus genotype (MLG). A total of 129 different MLGs were obtained. In the pea race, 195 out of the 559 parthenogenetic females collected established a colony (here again because of the high prevalence of parasitism). Genotyping at the 15 microsatellite loci revealed that there were only 88 different MLGs.

Then, we extended our investigations to another aphid species (*M. persicae*) by reanalyzing a large dataset of lineages that was collected in sympatry in France between 2011 and 2014 (Roy et al., 2013, 2022). A total of 1119 parthenogenetic females were collected, and genotyped at 14 polymorphic microsatellite loci (MP_27, MP_39, MP_44, MP_5, MP_7, MP_23, MP_45, MP_28, MP_9, MP_13, MP_2, MP_38, MP_4, MP_46) following DNA extraction and PCR protocols described in Roy et al. (2022). A total of 467 different MLG were identified using the same methods described above for *A. pisum*. We refer to this dataset as the “sympatric *M. persicae* dataset”.

Finally, we also reanalyzed another *M. persicae* dataset which consists of 127 different MLG samples collected from 19 countries collected on 14 host plants whose genomes have been individually sequenced using Illumina paired-end (Singh et al., 2021). We restricted our analyses to the 96 samples of *M. persicae sensu stricto*, by eliminating samples collected on tobacco, which likely belong to the subspecies *M. persicae subsp. nicotianae* (Blackman, 1987; Margaritopoulos et al., 2007). We started our analyses from the vcf file produced by Singh et al. (2021). We refer to this dataset as the “worldwide *M. persicae* dataset”.

Inference of reproductive mode

We inferred the mode of reproduction for each of the aforementioned MLG lineages of *A. pisum* and *M. persicae* in different ways, depending on available information and data. In the *A. pisum* alfalfa race, the mode of reproduction was inferred from the geographical origin of lineages, assuming that only CP lineages can survive in cold-winter regions (thanks to cold-resistant eggs), and that OP lineages are favored in regions with mild-winter regions. This is a rather crude way of inferring the mode of reproduction, but these inferences have been corroborated by other analyses (Jaquiéry et al. 2014).

For the *A. pisum* pea and clover races, the reproductive mode was determined in the laboratory, by assessing for each lineage whether it produced sexual females when exposed to sex-inducing conditions (as described in Le Trionnaire et al., 2009). Briefly, one one-day-old larva was placed on a faba bean plant *Vicia faba*, the universal host plant for *A. pisum*, kept at 18°C with a 16:8 L:D photoperiod for one week and then transferred to short-day conditions (12:12 L:D regime, 18°C). Six days later, one of its offspring (i.e., the second generation) was collected, placed on a new plant and maintained under short-day conditions. When it reached adulthood, this second-generation individual produces first females (oviparous sexual females for CP lineages, parthenogenetic females for OP lineages), then males and finally parthenogenetic females. The second-generation individual was transferred to a new plant each week (to avoid overcrowding due to the large number of offspring). The offspring were kept at 18°C under a 16:8 L:D photoperiod and their reproductive status (sexual female, male or parthenogenetic female) was determined individually on the basis of morphological characteristics (Miyazaki, 1987) when they were adults. The capacity of a lineage to produce winged males was also recorded. Three replicates were performed per lineage. We found that out of the 99 clover lineages tested (some died in the meanwhile), 76 lineages produced sexual females (hence were assigned as CP), and 23 did not (hence were assigned as OP). We observed a clear geographical effect, since the 23 OP lineages all originated from western France, while all but one of the 76 CP lineages originated from eastern France. In the pea host race, in which we were able to measure the mode of reproduction in the 61 lineages that survived until phenotype assessment, we identified 17 OP lineages and 44 CP lineages, but the reproductive mode was less associated with geographical origin.

In *M. persicae*, the reproductive mode was inferred from the plant on which the parthenogenetic female was sampled. In fact, this species alternates between two distinct types of

host plants in order to complete a full annual cycle of cyclical parthenogenesis. The part of the cycle involving sexual reproduction (which includes the sexual morphs, the parthenogenetic females that hatch from the sexually produced eggs and the 1 to 3 subsequent parthenogenetic generations) shows a very strong host specialization, and exclusively occurs on the peach tree *Prunus persica* (Blackman, 1974). In contrast, the major part of the asexual phase takes place on an exceptionally diverse range of host plants (more than 400 species from 40 different plant families, Blackman & Eastop, 2000), but never on peach trees. For this reason, we have assumed that the 382 MLGs collected on peach trees and represented by only one aphid sample have the CP phenotype in the sympatric *M. persicae* dataset (Roy et al., 2013, 2022). However, aphids collected on secondary hosts could be OP or CP, so we considered as OP only the 85 MLG found exclusively on secondary hosts (in order to avoid falsely including CP in this group as much as possible). For the worldwide *M. persicae* dataset (Singh et al. 2021), we only used the host plant from which the clones were collected to infer the mode of reproduction, as no information on MLG redundancy was available.

As we used different biological criteria to infer the mode of reproduction, which may lead to different levels of precision, we tested whether the genetic characteristics of the populations inferred as OP and CP matched theoretical predictions. In particular, the ratio of genotypic diversity G to the number of samples N (i.e. G/N) should be close to 1 for a CP population (because of regular sexual reproduction), but much smaller for an OP population (Halkett et al., 2005). For each dataset, we tested whether the G/N differed between populations made of OP and CP lineages using a two proportion Z-test. Second, clonal divergence through the independent accumulation of new mutations in OP genomes should also lead to a particular configuration of OP lineages in the phylogenetic tree. These lineages are expected to be scattered at the end of short branches in the tree, resulting in a twiggy phylogenetic distribution of OP lineages. In contrast, recombination and mixing of genes in CP lineages causes greater and more uniform genetic distances among lineages (Barraclough et al., 2003; Higgs & Derrida, 1992; Tang et al., 2014). To investigate these patterns, the Prevosti distance (Prevosti et al., 1975, *prevosti.dist* function of the R package *poppr*, Kamvar et al., 2014) was calculated between pairs of individuals for each dataset (the three *A. pisum* races and the two *M. persicae* dataset). The Prevosti distance was calculated using 14 to 15 microsatellite loci for the *A. pisum* host races (Suppl. Table 1) and for the sympatric *M. persicae* dataset (Suppl. Table 2). In both cases samples with missing data at three or more markers were eliminated. For the *M. persicae* worldwide dataset (for which individual resequenced genomes were available), the Prevosti distance was measured on 574,677 SNPs. Neighbor-joining trees were then drawn with the function *nj* from the *ape* package (Paradis & Schliep, 2019) and bootstrap values were computed over 1000 replications with the *aboot* function (package *poppr*). For each tree, the lengths of the terminal branches (from the tip to the nearest node) for each lineage were recovered (Suppl. Figures 1 and 2) and difference in branch lengths between OP and CP lineages was tested with two-sided Mann-Whitney tests.

Library preparation for pool-sequencing

The number of libraries for pool sequencing constructed for the different host races/species differed depending on opportunities and sample structure. In the *A. pisum* alfalfa race, the 3 OP populations (made of 14 different lineages each) and 3 CP populations (made of 20 to 21 lineages each) had already been illumina-sequenced in two replicates for a previous study (Rimbault et al. 2023), resulting in a total of 12 poolseq libraries (Table 1). For the *A. pisum* clover race, in which there was an almost perfect association between the inferred reproductive mode and geography, the 23 eastern OP lineages were used to form an OP library, and 23 and 22 randomly selected CP lineages (among the 75 eastern CP

lineages) were used to form two independent CP libraries, leading to a total of 3 libraries. In the *A. pisum* pea race, we constructed a total of 6 libraries to keep track of the geographical origin of the lineages (OP and CP lineages occurred in the two regions). The number of individuals per library was variable (ranging from 3 to 28) due to the small number of MLGs available for some conditions. Finally, for the sympatric *M. persicae* dataset, a subset of ~77 MLG out of the 467 MLGs were selected on the basis of their inferred reproductive mode and the availability of DNA extracts (obtained in previous studies, Roy et al. 2013, 2022). The neighbor-joining tree based on the Prevosti genetic distance revealed a structuration of these 77 MLGs (see the Results section and Suppl. Fig. 1). A total of 22 CP lineages formed a clade clearly separated from the other 18 CP and 37 OP lineages that were intermingled. To identify the region(s) of the genome associated with sex loss, CP lineages with reduced overall differentiation from OP lineages are of great interest because they could minimize genome scan noise related to other potential differences between the two types of lineages. On the other hand, the clearly separated cluster of 22 CP lineages may confirm (or not) the candidate regions identified with the more related CP lineages. We therefore formed two CP libraries, one made of the 22 well-differentiated CP lineages (the MP_CP1 pool) and a second with the 18 remaining CP lineages (the MP_CP2 pool). To have a similar number of lineages for the OP libraries, we separated the 37 OP lineages in two groups. The 18 OP MLGs found multiple times over several years, strongly suggesting obligate parthenogenetic reproduction were grouped to form one library (the MP_OP1 pool) and the 19 OP MLGs observed only once during the 4 years of fieldwork were grouped into a second library (the MP_OP2 pool).

To prepare the different libraries, DNA was extracted from one adult parthenogenetic female per lineage using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer instructions. After RNase treatment, DNA solutions were pooled in equimolar proportions for each population ensuring that each lineage contributed equivalent amounts of DNA to the pool. Paired-end 150 bp libraries were constructed per population from these DNA pools using the Genomic DNA Sample Preparation Kit (Illumina, San Diego, CA). The resulting libraries were sequenced on Illumina NovaSeq using S4 flowcells chemistry at the GetPlage Platform of Genotool-Toulouse (for the *A. pisum* samples) and at the University of Exeter (for the *M. persicae* samples). The raw data will be publicly available at the Sequence Read Archive of the NCBI database soon.

Identification of SNPs with pool-seq data

We used the JIC1 reference genome of the pea aphid (available on AphidBase: *Acyrtosiphon pisum* JIC1 genome v1.0, (Mathers et al., 2021), which is the most accurate genome version for this species. Paired-end reads were mapped to the fasta file of the genome assembly. For the *A. pisum* alfalfa race, the alignment (BAM) files corresponding to the 6 OP and 6 CP libraries were merged in one mpileup file using samtools (Danecek et al., 2021). We created a sync file using Popoolation2 (Kofler et al., 2011) with default parameters except for a minimum base quality set to 20. We then used additional filters to retain only reliable and informative SNPs for genome scan analysis. First, only biallelic SNPs for which each allele was carried by at least four reads considering all libraries were kept. In the case where 3 alleles were found at one position, one of which represents a single read or a deletion, the latter was ignored and the SNP was considered biallelic. We applied further filters on sequencing depth to avoid duplicated genomic regions and to remove SNPs with non-reliable allele frequency estimates. Only SNPs with a sequencing depth higher than 5% and lower than 95% of the total sequencing depth per library were considered. A minor allele frequency (MAF) threshold of 5% was applied to eliminate SNPs harboring rare alleles and which are not informative for a genome scan. Finally, unless specified, all

analyses were performed on the summed allele counts for CP libraries on one hand and for OP libraries on the other hand, to have only two pools.

We then re-used exactly the same procedure for variant calling on the three libraries of the *A. pisum* clover race (one OP and two CP) and on the four libraries of *M. persicae* (two OP and two CP) to end-up with two pools per species (except that we used the *Myzus persicae* G006 reference genome v3.0 for *M. persicae* libraries, Singh et al., 2021). In the pea host race of *A. pisum*, the large difference in the number of individuals among the six libraries (two OP and four CP libraries, with 3 to 28 lineages per library, see Suppl. Table 3) prevented to directly sum them. To give equal weight to each sequenced individual (regardless of whether it was in a library with a small or large number of individuals), the fastq files were subsampled to similar total number of sequenced reads per individual and pooled per reproductive mode to form a single OP pool and a single CP pool. The same methods as described above were then used for variant calling on these two artificially formed libraries for the pea race. Overall, the average sequencing depth per pool ranged from 123.9 to 15.1, and we obtained datasets of 3,657,027 SNPs for the *A. pisum* alfalfa race, 3,552,888 SNPs for the *A. pisum* clover race, 2,785,358 SNPs for the *A. pisum* pea race and 1,164,226 SNPs for the sympatric *M. persicae* dataset.

Finally, in order to study the genetic relationships of lineages from the different *A. pisum* races, we also created a SNP dataset following the same steps as described above, but considering the three races altogether so as not to miss any host race-specific SNPs that could be informative in a phylogenetic context. The final dataset contained 3,270,482 SNPs.

Detection of genomic regions associated with reproductive mode in pool-seq data

To scan for associations between allele frequencies and reproductive mode variation, the F_{ST} at each SNP between the OP and CP pools of the *A. pisum* alfalfa race were computed using the R package poolstat (Hivert et al., 2018). We then calculated the average F_{ST} within 100-kb windows sliding by 10-kb steps to smooth its variation along the genome. The genomic regions of increased genetic differentiation between reproductive modes, that would score high values of F_{ST} , would be candidate region(s) for controlling the reproductive mode variation. Second, we also ran BayPass on the two pools under the core model using the $C2$ statistic (Olazcuaga et al., 2020). This method aims at identifying SNPs associated with a binary trait (here OP or CP reproductive mode) by comparing the standardized population allele frequencies (i.e., the allele frequencies corrected for the population structure). The F_{ST} and the $C2$ indices were calculated in the same way for the *A. pisum* clover and pea races and for the sympatric *M. persicae* dataset. The regions showing the highest F_{ST} or $C2$ values were retained as candidate regions.

Detection of genomic regions associated with reproductive mode in individual genomes

From the vcf file of individually resequenced genomes of *M. persicae* collected worldwide (Singh et al. 2021), we kept the 50 presumed CP and 46 presumed OP lineages to estimate allele frequencies. A total of 1,213,868 SNPs exceeded the minor allele frequency threshold ($MAF > 5\%$) and were therefore retained. First, to facilitate the comparison with the Pool-seq results, we applied the same method (F_{ST}) to the individual genome dataset. The F_{ST} per SNP was calculated as above on the summed allele counts for the CP lineages on one the hand and the OP lineages on the other hand, and then averaged over 100-kb windows sliding by 10-kb steps. Second, individually resequenced genomes enable the use of more efficient methods for GWAS, particularly those based on phased haplotype size and frequency, such as the XPEHH approach (Cross population extended haplotype homozygosity, Sabeti et al., 2007), implemented in the selscan v1.2.0 program (<https://github.com/szpiech/selscan>, Szpiech &

Hernandez, 2014). We performed haplotype phasing with SHAPEIT2.v904 (Delaneau et al., 2012) with default parameters and then computed XPEHH score for each chromosome, before normalization in R. Interestingly, XPEHH can identify which of the two compared populations underwent a selective sweep, depending on which population is specified as the ancestral population for the trait under investigation. We therefore performed the XPEHH tests in both directions to identify selective sweeps that occurred specifically in the expected (OP) population (since OP is the derived trait in aphids, Davis, 2012).

Characterization and comparison of candidate regions identified in the different species

A few well supported candidate regions were identified in some *A. pisum* races and in *M. persicae*. First, we looked at the putative functions of the genes located within candidate regions, and also tested, when applicable, for GO terms enrichment against the whole set of genes with topGO R package (v2.48.0, Alexa & Rahnenfuhrer, 2023) and rrvgo R package (v1.12.0, Sayols, 2023) using Fisher's exact tests ($p < 0.01$). Second, we tested whether the candidate regions identified in *A. pisum* and *M. persicae* could be orthologous. For this, we identified syntenic blocks of genes between the reference genomes of *M. persicae* and *A. pisum* using MCScanX v1.1 (Wang et al., 2012). We first carried out a blastp search of protein sequences using blast (version blast+ 2.5.0) and then ran MCScanX with default parameters. MCScanX results were visualized with SynVisio (<https://synvisio.github.io/#/>, last accessed August 1, 2023). Third, even if the candidate regions are not globally orthologous, one or more genes may be (if movement of individual genes has occurred). This was also investigated with the previous analyses. Finally, we also tested whether some enriched GO terms were shared between candidate regions, which could indicate similar functional roles.

For the worldwide *M. persicae* dataset (which included individual genome sequences), we drew a phylogenetic tree based on the candidate region (using the methods described above) and compared it to the tree based on the whole genome. This allowed us to assess whether the distribution of asexual and sexual lineages was consistent with the fact that the region on which the tree is built is a potential candidate for control of reproductive mode variation. For the *A. pisum* races, we also compared how the different OP and CP pools from the different host races grouped in trees constructed using only the region corresponding to the main *A. pisum* alfalfa candidate regions or the whole genome.

RESULTS

Genotypic characteristics of OP and CP lineages in A. pisum and M. persicae

The number of lineages inferred as OP varied from 17 to 85, and that inferred as CP from 43 to 382 depending on the dataset (i.e. three *A. pisum* races and the sympatric and worldwide *M. persicae* datasets, Table 1). We then checked whether the populations made of OP or CP lineages actually showed the genetic signatures expected under their respective modes of reproduction (Halkett et al. 2005). First, we tested whether the genotypic redundancy (measured by the ratio of the number of different MLGs to the number of samples, i.e. G/N) differed between OP and CP populations. This was the case for three out of the four datasets on which we could measure it. In these three datasets (the *A. pisum* alfalfa and clover races, the sympatric *M. persicae*), the G/N ratio was close to 1 in the CP

populations (≥ 0.89), as expected under sexual reproduction which creates an infinite number of unique genotypes. In contrast, this ratio ranged from 0.12 to 0.24 in the OP populations, supporting for MLG redundancy as expected under clonal reproduction. Only the *A. pisum* pea race did not match the predictions, showing similar G/N in both population types (Table 1). Second, clonal divergence through the independent accumulation of mutations in asexual lineages is predicted to result in a twiggy phylogenetic distribution, with asexual lineages appearing at the end of short terminal branches (Barraclough et al. 2003). In each of the trees (obtained with 14-15 microsatellites or 500,000 SNPs, Suppl. Figures 1-2) corresponding to the 5 datasets, we observed shorter terminal branches for OP than for CP lineages, the effect being highly significant in all cases ($p < 0.001$, Table 1). Overall, these results suggest that the inference of reproductive mode is generally reliable for conducting our GWAS-type approach.

No evidence for a shared genomic region associated with sex loss in A. pisum host races

The genome-wide F_{ST} analyses revealed a stronger differentiation between the OP and CP populations of *A. pisum* clover race (F_{ST} was 0.064) and alfalfa race (0.042) than between those of pea race (0.016).

In the *A. pisum* alfalfa race, the re-analysis of the Rimbault et al. (2023) dataset revealed one main candidate region, characterized both by F_{ST} values greater than 0.4 and by BayPass $C2$ values above 30 (Figure 1A and 1B), which corresponds to the region identified previously on an earlier genome version. On the JIC1 reference genome we used here, this 840 kb region locates between positions 84,765,000 and 85,605,000 on the X chromosome (Fig. 2A) and contains 42 predicted genes (Suppl. Table 5), instead of the 32 on the earlier genome version used by Rimbault et al. (2023). Half of these 42 genes have annotated functions (see Suppl. Table 5), and enriched GO terms included the regulation of transcription by RNA polymerase II, RNA metabolic process and tetrahydrofolate biosynthetic process.

In contrast, the differentiation patterns observed in the clover and pea races differed greatly from the one found in the alfalfa race, with, in both cases, several regions that stand out moderately and are not necessarily supported by the two indices (the F_{ST} and $C2$). In the *A. pisum* clover race, three regions showed F_{ST} higher than 0.4, and five additional ones above 0.3 (Fig. 1C). However, only one of these eight regions (the one on the X chromosome with the highest F_{ST} values, ~ 0.44) was also well supported by the BayPass $C2$ statistic (Fig. 1D) (Fig. 2B). This region appeared to be close to the candidate region identified in the alfalfa race (Figure 1A), but fine inspection revealed that the clover candidate region locates at 100-300 kb from the alfalfa 840 kb region (Figure 2B). Interestingly, this region corresponded to a previously described locus (called *Aphicarus*) controlling the wing dimorphism in pea aphid males (Caillaud et al., 2002; Li et al., 2020), which harbors two alleles (the winged and the wingless allele). The phenotype of males is determined by the allele carried on their single X chromosome (males are XO). The genotype of a parthenogenetic female lineage (which carries two X copies) at *Aphicarus* can therefore be inferred as homozygous for one of the two alleles - if all the males produced by the lineage show the same phenotype - or heterozygous - if both types of males are produced. We checked whether our observations of the frequency of winged males in the lineages used in the poolseq were compatible with this strong differentiation at *Aphicarus*. In the CP lineages, only one lineage out of the 36 that produced males was heterozygous for the winged allele, the frequency of the winged allele was thus low ($1 \cdot 0.5 / 36 = 0.013$). The winged allele appeared to be more frequent in the OP lineages, with presumably 2 homozygous and 2 heterozygous lineages in the 14 lineages that produced males (hence a frequency of $(2 \cdot 1 + 2 \cdot 0.5) / 14 = 0.21$).

Table 1. Genetic characteristics of OP and CP populations for three host races of the pea aphid complex (*Acyrtosiphon pisum*) and for *Myzus persicae* (sampled in sympatry in France or worldwide). The number of individuals used in GWAS approaches is also indicated, as well as the source of the datasets.

	<i>Acyrtosiphon pisum</i>						<i>Myzus persicae</i>			
	Alfalfa race		Clover race		Pea race		Sympatric French populations		Worldwide populations	
	Pool-sequencing		Pool-sequencing		Pool-sequencing		Pool-sequencing		Individual genome resequencing	
Inferred reproductive mode ^a	OP	CP	OP	CP	OP	CP	OP	CP	OP	CP
Sample size (N)	320	197	94	83	43	122	690	429	na	na
No. of genotypes (G)	72	179	23	76	17	43	85	382	na	na
G/N^b	0.23	0.91 ^{***}	0.24	0.92 ^{***}	0.40	0.35 ^{ns}	0.12	0.89 ^{***}	na	na
Terminal branch length^b	0.14	0.19 ^{***}	0.15	0.23 ^{***}	0.15	0.21 ^{***}	0.06	0.13 ^{***}	0.036	0.064 ^{***}
No. of poolseq libraries for GWAS	6	6	1	2	2	4	2	2	na	na
No. of lineages used in GWAS	42	61	23	45	17	44 ^c	37	40	46	50
Origin of the data set	Jaquiéry et al., 2014, Rimbault et al., 2023		Defendini et al., 2023		This study		Roy et al., 2013, 2022		Singh et al., 2021	

^a The reproductive mode (OP or CP) was inferred using different approaches depending on aphid taxa. The geographical origin of samples was used to infer reproductive mode in the alfalfa race of *A. pisum*, reasoning that OP lineages are favored in mild-winter regions and CP in cold-winter regions (Jaquiéry et al., 2014). For the clover and pea races of this complex, we assessed the phenotype of individuals produced under conditions that normally induce the production of sexual morphs in the pea aphid (Le Trionnaire et al., 2008), CP lineages being able to produce sexual females, but not OP lineages. The reproductive mode in *M. persicae* was inferred from the host plant on which the parthenogenetic female was collected, assuming that only CP lineages occur on the primary host plant (on which the sexual phase takes place) while lineages collected on the secondary host plants are likely to be OP (and to some extent CP). Reproductive mode inference was globally verified with two indices for which theory predicts differences between asexual and sexual populations: genotypic diversity (G/N) and length of terminal branch retrieved from neighbor-joining trees (built using Prevosti genetic distances measured on samples used in the genome scan analyses).

^b The significance of the differences in terminal branch length (respectively, G/N) between the OP lineages and the CP lineages within the different host races or species was obtained using the two-sided Mann-Whitney tests (respectively, two-sided proportion Z-tests). ns: p>0.05, *: p<0.05, **: p<0.01, ***: p<0.001.

^c Another CP lineage collected on the pea host plant in western France (clone P123), which was maintained in the laboratory, was included in the GWAS analysis in order to increase the sample size.

These allele frequencies were not sufficient to generate an $F_{ST} > 0.4$, but it should be noted that the wing phenotype remained unknown for 9 CP lineages (out of 45 in the pools) and 9 OP lineages (out of 23) that did not produce males in laboratory conditions. If the association between *Aphicarus* alleles and reproductive modes is high in these 18 lineages, then our data would be compatible with an F_{ST} of 0.4 on *Aphicarus*. Contrastingly, in the pea race, all males produced were winged in both OP and CP lineages, which is compatible with the low differentiation observed at *Aphicarus* in this race (Figure 2C). After discarding the *Aphicarus* region as candidate for sex loss in the clover race, the next two best candidate regions (characterized by windows with $F_{ST} > 0.4$) were a 300 kb region located around positions 26,000,000 on autosome A3 and an 870 kb region at position 36,500,00 on the X chromosome. However, these regions were moderately supported by the C2 statistic. GO term enrichment analyses of the 19 predicted genes in the A3 candidate region revealed an enrichment in receptor signaling pathway, cilium assembly, microtubule-based movement, carbohydrate metabolic process and intracellular transport, while the 59 genes in the candidate region on the X were enriched with GO terms involved in viral process and regulation of cell cycle (Suppl. Table 7). Overall, the large number of regions showing moderate differentiation between OP and CP lineages in the clover race makes it difficult to identify a candidate region with confidence and suggests that sex loss may be controlled by a combination of loci with low effects.

In the *A. pisum* pea race, no region stood out from the rest of the genome in the F_{ST} scan, the windows with the maximal F_{ST} value being less than 0.17, indicating a low level of association between the allelic variants and the reproductive phenotypes. We nevertheless observed a slight increase in F_{ST} in the middle of the X chromosome (Figure 1E, 2C) and the window with the highest F_{ST} value in the whole genome (0.167) was located just before the alfalfa 840-kb candidate region (Fig. 2C). This region was however not supported by the C2 statistic, which rather identified linked SNPs in other regions (Figure 1F). The five SNPs with the highest C2 values (above 40) points to four different genomic regions. The SNP with the highest C2 value (52) is located on the A2 chromosome (at position 24965). However, a careful investigation revealed that the beginning of the A2 autosome (from positions 0 to 33,000) corresponded to a part of the genome sequence of a facultative symbiont of the pea race, *Serratia symbiotica*. Visual inspection of the coverage of this region in the different pea pools with IGV-2.16.1 (Robinson et al., 2011) suggests that this is not a true nuclear insertion of the *Serratia symbiotica* genome, but an error in genome assembly. Indeed, coverage is anomalously low in all pea pools compared to expectation for nuclear regions. We therefore suspected that some reads from the *Serratia symbiotica* symbiont mapped on this poorly assembled part of the genome. However, the fact that one highly differentiated SNP was observed among OP and CP pools may indicate that the OP and CP lineages differ in their *Serratia* strains. Further analysis is required to test this hypothesis. In contrast, the four remaining SNPs with high C2 seemed to locate on the true nuclear genome of the pea aphid: two were found on A2 (at positions 23027808 and 23027818) and located within the coding sequence of gene g14329 for which no annotation is available. The next high C2 SNP located on the X chromosome (at position 170693875), in an intron of an uncharacterized gene (g11555). The fifth SNP located on the same chromosome at position 53327557, in a non-coding sequence. Overall, although further investigations are required in the pea race, these analyses reinforce the conclusions obtained with the clover race, namely that the genetic basis of sex loss is not shared between the three *A. pisum* races and that the genetic architecture of this trait seems more complex in the pea and clover races than in the alfalfa race.

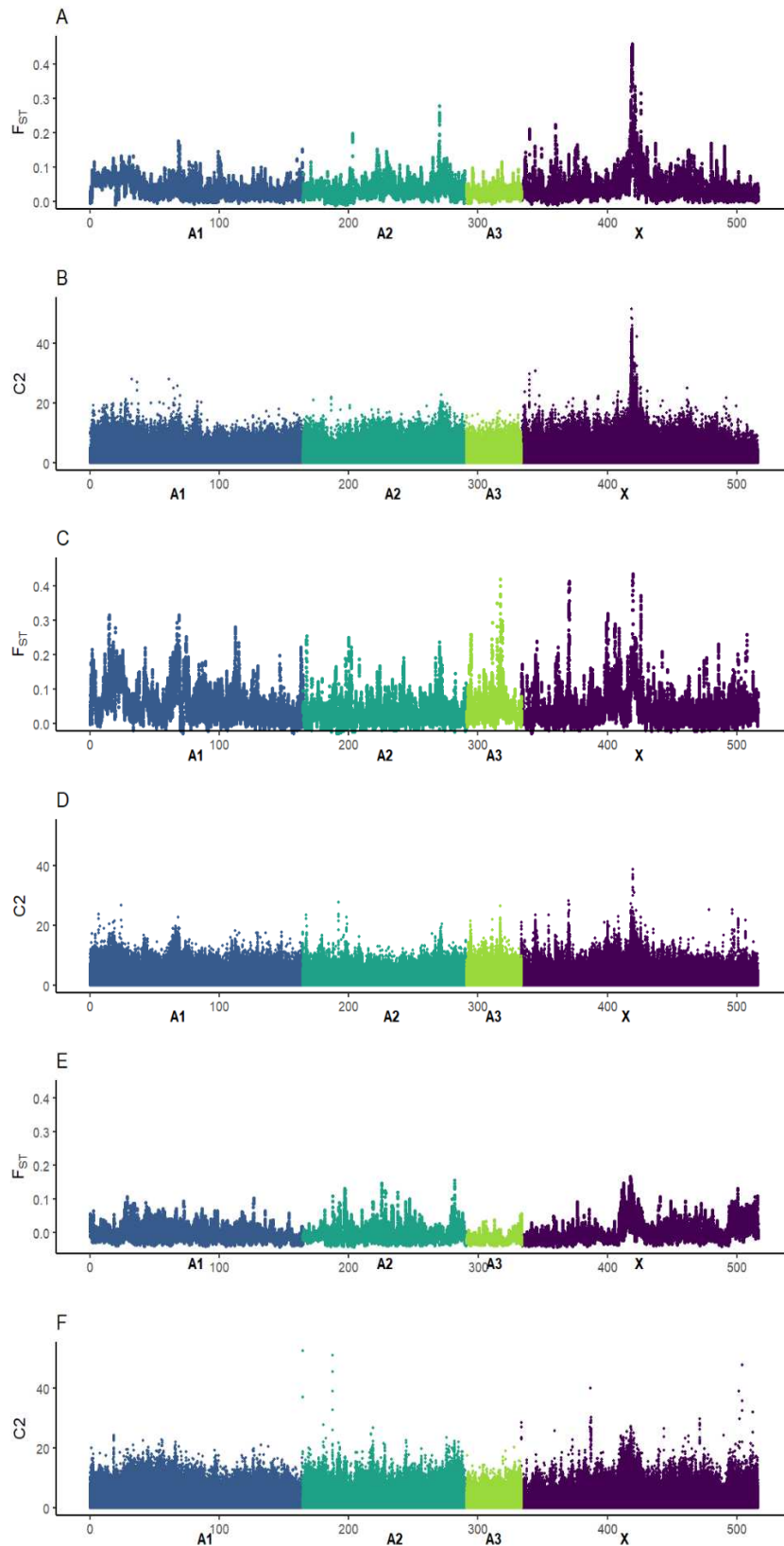


Figure 1. Characterization of genomic regions associated with sex loss in three host races of the pea aphid *Acyrtosiphon pisum*. Genetic differentiation (F_{ST}) is measured between obligate parthenogenetic (OP) and cyclical parthenogenetic (CP) populations of the alfalfa (A), clover (C) and pea races (E) in 100-kb windows sliding by 10-kb steps. The C2 statistic from the BayPass analysis between OP and CP populations is showed for each SNP for the alfalfa (B), clover (D) and pea (F) races.

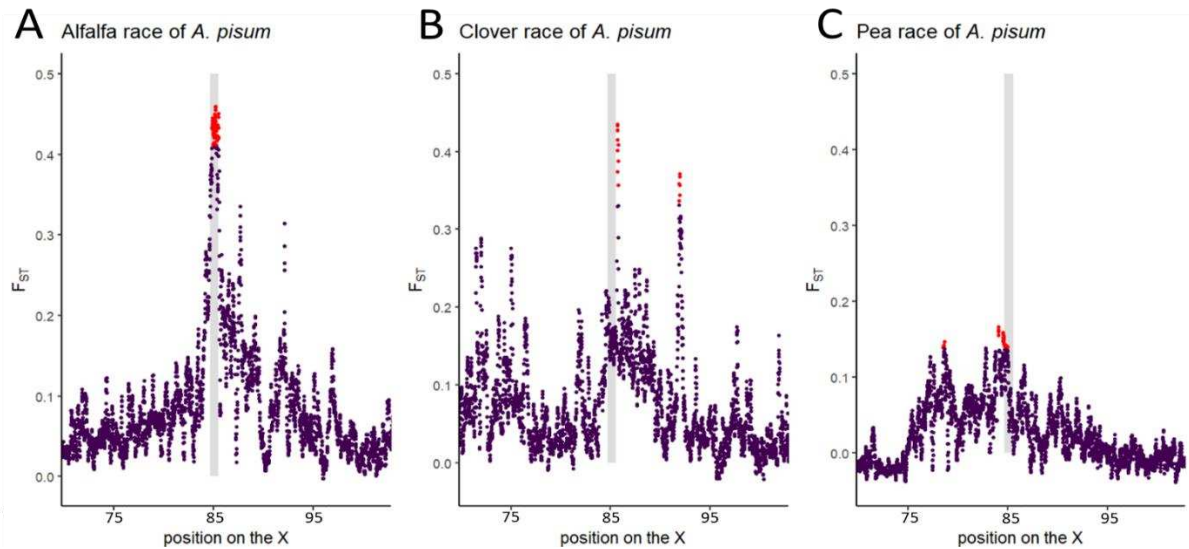


Figure 2. Genetic differentiation between obligate parthenogenetic (OP) and cyclical parthenogenetic (CP) populations of different *Acyrthosiphon pisum* host races for the part of the X chromosome carrying the 840 kb candidate region identified in the alfalfa host race (displayed in grey). Panel (A) shows F_{ST} for the alfalfa host race, while panels (B) and (C) show it for the clover and pea host races, respectively. Red dots correspond to windows along the genome with F_{ST} values above the 0.999 percentile of all F_{ST} values in the corresponding race. F_{ST} values were calculated over 100-kb windows sliding by 10-kb steps.

Finally, though the alfalfa candidate region was not identified as candidate in the two other *A. pisum* races, we were still interested to examine whether this genomic region showed a different evolutionary history than the rest of the genome. Interestingly, whereas the phylogenetic tree based on the whole genome systematically groups the samples by host race, the one based on the SNPs located in the alfalfa candidate region grouped the alfalfa OP samples with the pea samples (OP and CP, Figure 3). This suggests that an introgression of a DNA segment from the pea race (not necessarily involved in the OP phenotype in that race) could be associated with the loss of sex in the alfalfa race. This also merits further investigation.

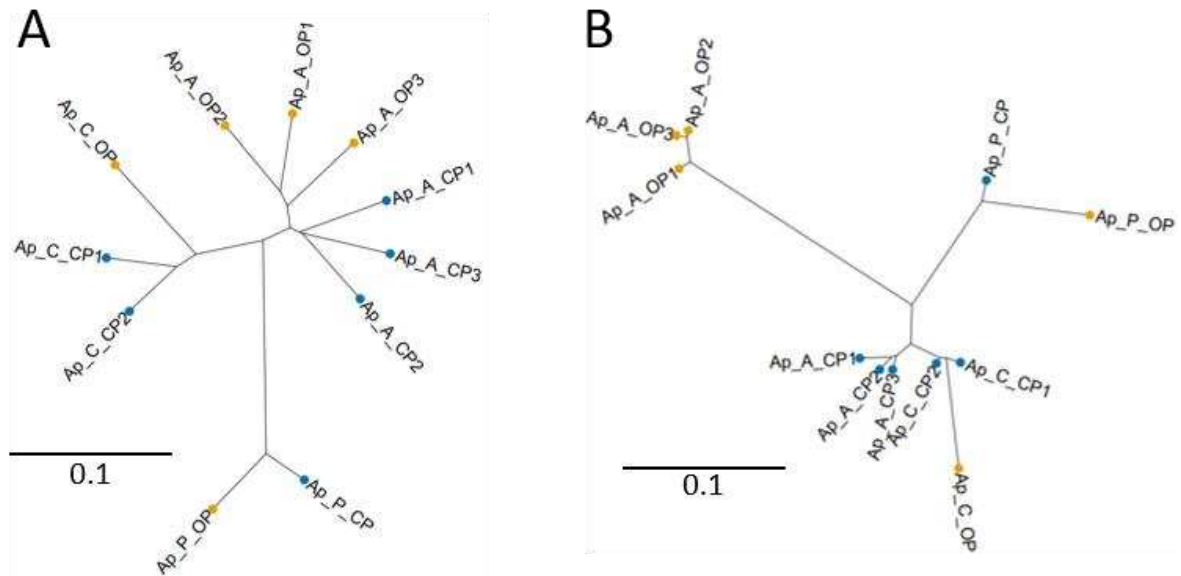


Figure 3. Genetic relationships between obligately parthenogenetic (OP) and cyclically parthenogenetic (CP) lineages of the alfalfa, pea and clover races of *Acyrthosiphon pisum*. Neighbor-joining tree based on the Prevosti genetic distance were constructed using the 11 pools across the whole genome (3,270,482 SNPs) **(A)** or using only the 6,822 SNPs from the genomic region identified as associated with sex loss in alfalfa race **(B)**. Pools of OP lineages are in orange and pools of CP lineages in blue. The letters in the pool names (A, C and P, respectively) represent the host races (alfalfa, clover and pea, respectively)

Several genomic regions are associated with reproductive mode variation in M. persicae

The different analyses of the sympatric and worldwide *M. persicae* datasets each identified at least three or more regions that may be associated with variation in reproductive mode (Figure 4). The F_{ST} approach carried out on the sympatric pool-sequenced OP and CP lineages revealed two major peaks (peaks 1 and 2, highlighted in grey in Figure 4A), as well as five other secondary peaks with slightly lower F_{ST} values. The BayPass C2 statistic supported peak 2, but not peak 1, and identified a third region (peak 3, Figure 4B). Out of the top 10 SNPs with the strongest support according to the C2 statistic, five located under peak 3 near two genes: a NADPH-cytochrome P450 reductase (g12838), and a predicted gene with uncharacterized function (g12837). One of these five SNPs located within the coding sequence of g12838. Four other SNPs from the top 10 located under peak 2 within non-coding sequences. The last SNP in the top 10 was located at the beginning of the A1 chromosome within a non-coding sequence and did not co-localize with any of the identified F_{ST} peaks.

Clearer results were obtained in the worldwide *M. persicae* dataset, where the F_{ST} analysis of individual genomes revealed three highly differentiated regions (peaks 1, 2 and 4 on Figure 4C). Interestingly, these three main regions were all supported by the XPEHH approach, which relies on a different statistic - haplotype length and frequency - to identify regions evolving under selection (Figure 4DE). Furthermore, two of them also corresponded to those already identified with the poolseq sympatric dataset (peaks 1 and 2, Figure 4AB). In the XPEHH analysis, peaks 2 and 4 were identified as having undergone selection in the CP lineages (Figure 4D). This allowed us to rule out their involvement in sex loss, the signal being expected solely in OP lineages since CP is the ancestral phenotype. Such regions that underwent selection in CP populations only could be associated with traits involved in the

CP life cycle (e.g. host alternation, adaptation to the primary host, see discussion). In contrast, the third peak identified with F_{ST} in Figure 4C (peak 1) showed a selection signature only in OP lineages with the XPEHH approach (Figure 4E), making it a good candidate for sex loss. XPEHH also identified another region in OP lineages (peak 5, Figure 4E), but it was not supported by the other approaches.

Overall, even if several candidate peaks were identified in *M. persicae*, peak 1 (from positions 11,680,068 to 11,789,976 on chromosome A3) was identified with the two independent datasets (the sympatric poolseq and the worldwide individual genome datasets) and with different approaches (namely F_{ST} and haplotype characteristics), making it the best candidate. There were 8 genes in the region under peak 1, with few annotated functions. GO terms indicated that some were involved in the glycosylation and ubiquitination of proteins. The gene content of peaks 2 to 5 (Figure 4) was also examined (Suppl. Table 9). GO terms indicated some enrichment in translation and regulation of transcription.

Finally, we examined whether the grouping of lineages by reproductive mode in trees was better when only 350 SNPs from the best candidate region (i.e. peak 1) were considered (Suppl. Figure 3) than when whole genome SNPs were considered (Suppl. Figure 1B), as one would expect if this region was the main one controlling reproductive mode. However, this was not the case, suggesting again a complex genetic architecture for this trait variation in *M. persicae*.

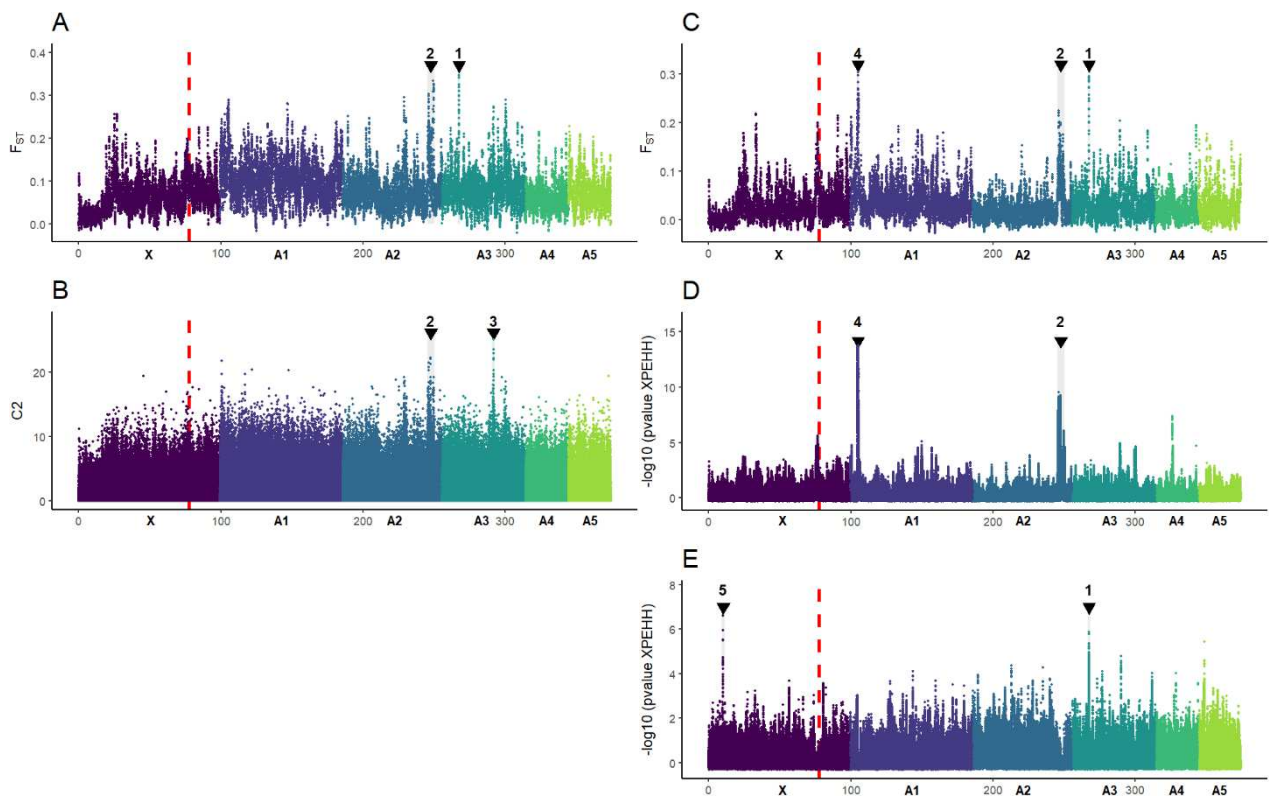


Figure 4. Detection of the genomic regions possibly associated with reproductive mode variation in *Myzus persicae* using two independent datasets (Panels A and B: the “sympatric dataset”, made of obligate parthenogenetic [OP] and cyclical parthenogenetic [CP] lineages sequenced in pools; panels C to E: the “worldwide dataset”, made of individually resequenced genomes of 96 OP and CP lineages). **(A)** Genetic differentiation (F_{ST}) between OP and CP pools (100-kb windows sliding by 10-kb steps). **(B)** C2 statistic from BayPass for each SNP between OP and CP pools. **(C)** Genetic differentiation (F_{ST}) between OP and CP populations measured from individually resequenced genomes (100-kb windows sliding by 10-kb steps). **(D)** $-\log_{10}$ transformed p -value from the XPEHH analysis (cross population extended haplotype homozygosity) for individually resequenced genomes of CP lineages and **(E)** same but for re-sequenced genomes of OP lineages. The numbers above the peaks help to identify the different regions discussed. The dashed red line corresponds to the *M. persicae* region orthologous the 840 kb candidate region for sex loss in the *Acyrtosiphon pisum* alfalfa race.

Candidate regions for sex loss are not shared between M. persicae and A. pisum

We first tested whether the candidate regions could be shared between the two species. To do this, we first restricted our investigations to the two regions for which support was strongest, namely the 840 kb region identified in the *A. pisum* alfalfa race and peak 1 in *M. persicae*. Half of the 42 genes from the 840 kb candidate region of the *A. pisum* alfalfa race have orthologs in the *M. persicae* genome (see Suppl. Table 6) and locate within an X-linked region that is conserved between the two species. This region (identified by a red dashed line on the *M. persicae* genome in Figure 4) has no homology with any of the *M. persicae* peaks identified with the different methods. We then tested whether some genes from these candidate regions could have the same predicted function, but this was not the case.

Moreover, these genes did not appear to be involved in similar biological processes: Among the 11 and 50 GO terms retrieved from the two candidate regions, only four were shared. These GO terms were involved in protein binding, oxidation-reduction processes or were associated with membrane components. Overall, the two candidate regions associated with sex loss in *M. persicae* and *A. pisum* are distinct, and we found no evidence for functional convergence between them. The other secondary regions identified as possible candidates in *A. pisum* (seven and three in the clover and pea races, respectively) and *M. persicae* (peaks 1 and 5) also showed no obvious signs of orthology. Nevertheless, we recognize that this analysis is incomplete (due to time constraints) and that the gene content of the secondary regions identified as potential candidates in *M. persicae* and *A. pisum* races need to be scrutinized in order to confirm that there is no common genetic basis for sex loss between these different taxa.

DISCUSSION

The aim of this work was to characterize the genetic basis of sex loss in a group in which transitions to obligate asexuality is common and in particular to test whether these transitions may share the same genomic architecture. To do so, we have worked along a continuum of divergence in aphid taxa, encompassing both intra- and inter-specific levels. We showed that the loss of sexual reproduction most likely involves different genomic regions even between closely related aphid taxa. Thus, this polymorphism in reproductive mode in the different taxa is not due to standing variation that has been maintained since their divergence, nor is the result of the spread of an *op* allele from one differentiated race to another (i.e., contagious asexuality) through hybridization. There was also no evidence for convergence neither in genomic localization nor in gene functions of the main candidate regions identified, suggesting that sex loss in aphids can be achieved in different ways. Finally, our analyses showed that there may be several loci associated with this trait in certain species.

In contrast to the unique and conspicuous candidate region responsible for sex loss in the alfalfa race of *A. pisum* (found by Jaquiéry et al., 2014 and Rimbault et al., 2023 and confirmed here), we found no such well-supported candidate region neither in *A. pisum* pea and clover races nor in *M. persicae*. In each of these aphid taxa, several regions with moderate to good support were found. In the clover race, once removed the locus involved in wing dimorphism in males (*Aphicarus*), seven genomic regions showed substantial genetic differentiation between OP and CP populations ($F_{ST} > 30\%$) and could be considered as candidates. Whereas no large genomic region stood out from the rest of the genome between the OP and CP lineages in the pea race based on 100 kb F_{ST} scan, the SNP-level analyses revealed three very localized regions (after excluding the *Serratia* misassembled region) potentially associated with sex loss. In *M. persicae*, which diverged from *A. pisum* around 22 million years ago (Mathers et al., 2021), a 109 kb genomic region containing 8 predicted genes was identified as the best candidate region because it was supported by different methods and datasets. However, a number of other regions stood out and cannot be totally discarded as candidates.

It is important to mention that the different regions showing increased differentiation between OP and CP lineages can be involved in the control of reproductive mode, but can also control other traits correlated or associated with this phenotype. This is all the more relevant as the ecology and habitat of the OP and CP lineages may differ (Blackman, 1974; Defendini et al., 2023; Frantz et al., 2006; Gilabert et al., 2009; Lehto & Haag, 2010). An excellent illustration of this is the clover race,

where the strongest signature of genetic differentiation between reproductive modes is observed at the *Aphidius* locus, probably due to a higher frequency of the winged allele in OP lineages than in CP lineages (an adaptive interpretation of this difference would be that the higher proportion of wing males in OP lineages would be selected for promoting mating with sexual females of CP lineages). The OP and CP lineages in the pea race could also carry different strains of *Serratia*, a facultative symbiont known to affect aphid thermal resistance (Pons et al., 2022; Tougeron & Iltis, 2022), generating a high C2 value for a SNP on the misassembled region of the genome, without this symbiont being the cause of the loss of sex (Simon et al., 2011). Similarly, in *M. persicae*, several differentiation peaks have been attributed to selection acting only on the CP lineages and could perhaps be related to the adaptation to the primary host (peach) or to host alternation, which are specific to the CP lineages (the OP lineages remain on the secondary hosts).

Crosses carried out between OP and CP lineages in *M. persicae* showed that two loci were necessary to explain the distribution of reproductive phenotypes observed in the offspring (Blackman, 1972), as for *S. avenae*, another aphid species (Dedryver et al., 2013). In the present study, we were not able to estimate how many loci are involved in determining sex loss in the pea and clover races of *A. pisum* and in *M. persicae* but we could unambiguously show that different genomic regions are involved in sex loss - as none of the putative associated regions are shared within the *A. pisum* complex or between species. This last result also indicates that independent losses have occurred and that they are relatively frequent, given that there is no shared basis even within the recently diverged *A. pisum* complex. Indeed, the divergence time between *A. pisum* races is estimated at 18,000-47,000 years using divergence of the maternally inherited obligate endosymbiont *Buchnera aphidicola* in host aphid lineages (Peccoud et al., 2009) to 419,000-772,000 years using nuclear divergence (Fazalova & Nevado, 2020). However, as the three races studied are known to be much closer phylogenetically to each other than to the radiation in *A. pisum* as a whole, the divergence of these three races should be more recent than these estimates. The diversity of mechanisms identified for sex loss in the aphid *Rhopalosiphum padi* - including spontaneous loss of sex (Delmotte et al., 2001), contagious asexuality by repeated gene flow from asexual lineages into sexual lineages (Delmotte et al., 2001; Simon et al., 1996) and hybridization (Delmotte et al., 2003) - is further evidence of the ease with which sex loss can occur in aphids.

These observations are in line with our understanding of how aphids can readily lose their ability to reproduce sexually. CP organisms such as aphids already have the genetic and developmental machinery required to perform parthenogenesis as it evolved in the early stage of aphid evolution. The transition to obligate asexuality therefore requires no novel adaptation, but only loss-of-function mutation(s) that will suppress the sexual part of the life cycle. In CP aphids, the production of the sexual morphs (i.e. sexual oviparous females and males) is induced by the shortening of the photoperiod in autumn. A loss-of-function mutation in any of the genes involved in the putative molecular cascade of sexual female production, from the early steps of photoperiod signal perception and transduction to the late steps of embryogenesis switch (from a parthenogenetic female to a sexual female phenotype) is theoretically sufficient to cause obligate asexuality. The number of genes involved in the whole process is likely to be numerous, as suggested by comparative transcriptomic studies of the establishment of sexual or parthenogenetic female phenotypes (Le Trionnaire et al., 2008, 2012; Huguet et al., *submitted*), but it can be also limited by the possible deleterious pleiotropic effects of certain loss-of-function mutations.

The ease with which loss of sex can arise in aphids contrasts with what is expected in obligate sexual taxa. Indeed, the evolution of asexuality in these taxa requires various cytological mechanisms

and changes to produce diploid eggs without fertilization (see Engelstädter, 2008; Schwander & Crespi, 2009). In particular, the ploidy level must be maintained and sometimes various processes such as sperm-dependent egg activation and paternal inheritance of centrioles must be bypassed. It is therefore likely that the transition to asexual reproduction requires a combination of different mechanisms, and that the asexual lineages observed in nature have undergone a consequent selective filter. The only other taxa where repeated transitions to asexuality has been studied is *Timema* stick insects. In this group, at least five independent transitions to obligate asexuality have been characterized, with the existence of closely related pairs of species that differ in their mode of reproduction (Schwander et al., 2011; Schwander & Crespi, 2009). Although the genetic basis of these transitions has not been elucidated, transcriptomic comparison of these pairs of species of different reproductive mode show some functional convergence (Parker et al., 2019), notably in genes involved in meiotic spindle formation and centrosome organization. These results suggest that the evolutionary path to asexuality is constrained in *Timema* insects, requiring changes of the same key genes. It remains unclear whether the frequent events of sex loss in this group are caused by an ancestral polymorphism (spontaneous parthenogenesis is observed in 1 to 2% of unfertilized eggs even in sexual *Timema* species) or are due to independent mechanisms. In any case, the absence of paternal inheritance of centrioles in all phasmids, coupled to the activation of the eggs by their passage through the reproductive tract in insects, facilitates transitions towards obligatory asexuality (Parker et al. 2018). So, like aphids, this group would be particularly predisposed to evolve towards asexuality due to the relative ease of acquiring this trait.

Our analysis in aphids did not reveal specific regions more likely involved in transitions to asexuality (none of the regions being shared between the *A. pisum* host races, nor between *A. pisum* and *M. persicae*). Nor did we observe any shared genes or convergence in the functional role of the candidate regions. However, we recognize that this latter analysis has some limitations, as a large proportion of the genes predicted in aphids do not have a characterized function. Nevertheless, comparison of the phylogenetic tree constructed from the whole genome and that based solely on the region corresponding to the 840 kb alfalfa candidate region revealed some interesting patterns. While the samples are grouped by host race in the whole genome tree, the history of the region linked to reproductive mode variation in the alfalfa race of *A. pisum* is strikingly different: Alfalfa OP populations cluster with pea samples, whether OP or CP, suggesting that the alfalfa *op* allele would come from introgression from the pea race (or another race closer to the pea race than the alfalfa race, Peccoud et al., 2009). This does not necessarily mean that the introgressed segment coded for an OP phenotype in its ancestral race. The introgression could simply have disrupted an interaction between genes in the molecular cascade involved in the production of sexual females, by bringing in genes with slightly different sequences or different regulatory sequences. This incompatibility could have resulted in an OP phenotype only in the latter race. These results merit further exploration, in particular a clear delimitation of the introgressed region (even if this probably corresponds exactly to the 840 kb identified) and careful investigation of its recombination patterns and evolutionary history when considering the 15 host races of the pea aphid complex. This possible history of hybridization could also explain the high divergence observed between the *cp* and *op* alleles in the alfalfa race, roughly estimated at 180,000-870,000 years ago (Rimbault et al., 2023). Given that new mutations leading to sex loss seem to arise frequently, it is surprising that the reproductive polymorphism in this region has been preserved for so long, without new *op* alleles evolving in other regions in the alfalfa race. If the *op* allele emerged by hybridization, then this reproductive polymorphism may be much younger than the estimated age of divergence between the alleles, resolving the apparent paradox.

Frequent *de novo* evolution of sex loss have important consequences on the level of genetic diversity of asexual lineages, leading to a pool of genetically diverse and polyphyletic asexual lineages (Simon et al., 2003). This source of genotypic variation can contribute to increase the adaptive potential of asexual populations (Janko et al., 2008; Lynch & Gabriel, 1983; Wright, 1977, reviewed in Neiman & Linksvayer, 2006), their competitiveness with sexual populations (reviewed in Vrijenhoek & Parker, 2009) and thus their persistence. Interestingly, most asexual taxa studied with sufficient detail are polyphyletic at the genus or even species level (Delmotte et al., 2001, 2003; Hebert et al., 1989; Johnson & Bragg, 1999; Schwander & Crespi, 2009).

In conclusion, this work provides a first overview of the evolutionary history of sex loss in aphids and strongly suggests that transitions to obligatory parthenogenesis, which are widespread in this group involved different genomic loci. Our results thus shed light on the reasons why sexual loss is ubiquitous in this group. Further analyzes are needed to support our conclusions. The minor peaks of genomic differentiation between OP and CP lineages and the gene contents under these peaks should be scrutinized with more attention to better support the absence of genetic or functional convergence among aphid OP lineages. QTL approaches would also be of great interest to firmly identify candidate regions, as what has been done for the alfalfa race (Jaquiéry et al., 2014). It will be also interesting to explore the difference in symbiont sequences between OP and CP lineages of the pea race of *A. pisum*, and test whether it is indeed due to strain difference. Finally, the multiple and independent events of evolution of sex loss at the inter- and intra-specific levels offer the opportunity to investigate the evolutionary consequences of sex loss. They would allow testing whether repeated losses of sexual reproduction lead to convergent changes in genomic sequences or gene expression patterns (e.g. Jaron et al., 2022; Parker et al., 2019).

ACKNOWLEDGMENTS

We thank the Get-Plage platform in Genotoul (Toulouse) as well as that of the University of Exeter for the sequencing of the DNA libraries. We thank Erick Cordeiro for the data transfer. This work was supported by grants from the French Research Agency (SexAphid ANR-09-GENM-017-001 and Mecadapt ANR-18-CE02-0012), the INRAE-SPE Department (AAP GenAsex and half a PhD grant for H.D.), the ERC (to C.B), and Région Bretagne (ARED, half a PhD grant for H.D.).

REFERENCES

- Alexa, A., & Rahnenfuhrer, J. (2023). *TopGO: Enrichment Analysis for Gene Ontology* (2.52.0). <https://bioconductor.org/packages/topGO>
- Bailleul, D., Stoeckel, S., & Arnaud-Haond, S. (2016). RClone: A package to identify MultiLocus Clonal Lineages and handle clonal data sets in R. *Methods in Ecology and Evolution*, 7(8), 966–970. <https://doi.org/10.1111/2041-210X.12550>
- Barraclough, T. G., Birky Jr, C. W., & Burt, A. (2003). Diversification in sexual and asexual organisms. *Evolution*, 57(9), 2166–2172.
- Barton, N. H., & Charlesworth, B. (1998). Why Sex and Recombination? *Science*, 281(5385), 1986–1990. <https://doi.org/10.1126/science.281.5385.1986>
- Bell, G. (1982). *The masterpiece of nature: The evolution and genetics of sexuality*. University of California Press.

- Blackman, R. L. (1971). Variation in the photoperiodic response within natural populations of *Myzus persicae* (Sulz.). *Bulletin of Entomological Research*, 60(4), 533–546. <https://doi.org/10.1017/S0007485300042292>
- Blackman, R. L. (1972). The inheritance of life-cycle differences in *Myzus persicae* (Sulz.) (Hem., Aphididae). *Cambridge University Press*, 62(2), 281–294. <https://doi.org/10.1017/S0007485300047726>
- Blackman, R. L. (1974). Life-cycle variation of *Myzus persicae* (Sulz.) (Hom., Aphididae) in different parts of the world, in relation to genotype and environment. *Bulletin of Entomological Research*, 63(4), 595–607. <https://doi.org/10.1017/S0007485300047830>
- Blackman, R. L. (1987). Morphological discrimination of a tobacco-feeding form from *Myzus persicae* (Sulzer) (Hemiptera: Aphididae), and a key to New World *Myzus* (*Nectarosiphon*) species. *Bulletin of Entomological Research*, 77(4), 713–730. <https://doi.org/10.1017/S0007485300012219>
- Blackman, R. L., & Eastop, V. F. (2000). Aphids on the world's crops: An identification and information guide. *Aphids on the World's Crops: An Identification and Information Guide.*, Ed. 2. <https://www.cabdirect.org/cabdirect/abstract/20001106823>
- Caillaud, M. C., Boutin, M., Braendle, C., & Simon, J.-C. (2002). A sex-linked locus controls wing polymorphism in males of the pea aphid, *Acyrtosiphon pisum* (Harris). *Heredity*, 89(5), 5. <https://doi.org/10.1038/sj.hdy.6800146>
- Caillaud, M. C., Mondor-Genson, G., Levine-Wilkinson, S., Mieuze, L., Frantz, A., Simon, J. C., & Coeur D'acier, A. (2004). Microsatellite DNA markers for the pea aphid *Acyrtosiphon pisum*. *Molecular Ecology Notes*, 4(3), 446–448. <https://doi.org/10.1111/j.1471-8286.2004.00676.x>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- Davis, G. K. (2012). Cyclical parthenogenesis and viviparity in aphids as evolutionary novelties. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 318(6), 448–459. <https://doi.org/10.1002/jez.b.22441>
- Dedryver, C. A., Le Gallic, J. F., Gauthier, J. P., & Simon, J. C. (1998). Life cycle of the cereal aphid *Sitobion avenae* F.: Polymorphism and comparison of life history traits associated with sexuality. *Ecological Entomology*, 23(2), 123–132.
- Dedryver, C.-A., Hullé, M., Le Gallic, J.-F., Caillaud, M. C., & Simon, J.-C. (2001). Coexistence in space and time of sexual and asexual populations of the cereal aphid *Sitobion avenae*. *Oecologia*, 128(3), 379–388. <https://doi.org/10.1007/s004420100674>
- Dedryver, C.-A., Le Gallic, J.-F., Mahéo, F., Simon, J.-C., & Dedryver, F. (2013). The genetics of obligate parthenogenesis in an aphid species and its consequences for the maintenance of alternative reproductive modes. *Heredity*, 110(1), 1. <https://doi.org/10.1038/hdy.2012.57>
- Defendini, H., Rimbault, M., Mahéo, F., Cloteau, R., Denis, G., Mieuze, L., Outreman, Y., Simon, J.-C., & Jaquier, J. (2023). Evolutionary consequences of loss of sexual reproduction on male-related traits in parthenogenetic lineages of the pea aphid. *Molecular Ecology*, 32(13), 3672–3685. <https://doi.org/10.1111/mec.16961>
- Delaneau, O., Marchini, J., & Zagury, J.-F. (2012). A linear complexity phasing method for thousands of genomes. *Nature Methods*, 9(2), 2. <https://doi.org/10.1038/nmeth.1785>
- Delmotte, F., Leterme, N., Bonhomme, J., Rispe, C., & Simon, J.-C. (2001). Multiple routes to asexuality in an aphid species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1483), 2291–2299. <https://doi.org/10.1098/rspb.2001.1778>
- Delmotte, F., Leterme, N., Gauthier, J.-P., Rispe, C., & Simon, J.-C. (2002). Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology*, 11(4), 711–723. <https://doi.org/10.1046/j.1365-294X.2002.01478.x>

- Delmotte, F., Sabater-Muñoz, B., Prunier-Leterme, N., Latorre, A., Sunnucks, P., Risper, C., & Simon, J.-C. (2003). Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. *Evolution*, 57(6), 1291–1303. <https://doi.org/10.1111/j.0014-3820.2003.tb00337.x>
- Dixon, A. F. G. (1998). *Aphid Ecology* (2nd ed.). Chapman & Hall.
- Engelstädter, J. (2008). Constraints on the evolution of asexual reproduction. *BioEssays*, 30(11–12), 1138–1150. <https://doi.org/10.1002/bies.20833>
- Fazalova, V., & Nevado, B. (2020). Low spontaneous mutation rate and Pleistocene radiation of pea aphids. *Molecular Biology and Evolution*, 37(7), 2045–2051. <https://doi.org/10.1093/molbev/msaa066>
- Frantz, A., Plantegenest, M., & Simon, J.-C. (2006). Temporal habitat variability and the maintenance of sex in host populations of the pea aphid. *Proceedings of the Royal Society B: Biological Sciences*, 273(1603), 2887–2891. <https://doi.org/10.1098/rspb.2006.3665>
- Gilabert, A., Simon, J.-C., Mieuze, L., Halkett, F., Stoeckel, S., Plantegenest, M., & Dedryver, C.-A. (2009). Climate and agricultural context shape reproductive mode variation in an aphid crop pest. *Molecular Ecology*, 18(14), 3050–3061. <https://doi.org/10.1111/j.1365-294X.2009.04250.x>
- Hales, D. F., Tomiuk, J., Woehrmann, K., & Sunnucks, P. (1997). Evolutionary and genetic aspects of aphid biology: A review. *EJE*, 94(1), 1–55.
- Halkett, F., Plantegenest, M., Bonhomme, J., & Simon, J.-C. (2008). Gene flow between sexual and facultatively asexual lineages of an aphid species and the maintenance of reproductive mode variation. *Molecular Ecology*, 17(12), 2998–3007. <https://doi.org/10.1111/j.1365-294X.2008.03798.x>
- Halkett, F., Simon, J.-C., & Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, 20(4), 194–201. <https://doi.org/10.1016/j.tree.2005.01.001>
- Hebert, P. D. N., Beaton, M. J., Schwartz, S. S., & Stanton, D. J. (1989). Polyphyletic origins of asexuality in *Daphnia pulex*. I. Breeding-system variation and levels of clonal diversity. *Evolution*, 43(5), 1004–1015. <https://doi.org/10.1111/j.1558-5646.1989.tb02546.x>
- Helden, A. J., & Dixon, A. F. G. (2002). Life-cycle variation in the aphid *Sitobion avenae*: Costs and benefits of male production. *Ecological Entomology*, 27(6), 692–701. <https://doi.org/10.1046/j.1365-2311.2002.00451.x>
- Higgs, P. G., & Derrida, B. (1992). Genetic distance and species formation in evolving populations. *Journal of Molecular Evolution*, 35(5), 454–465. <https://doi.org/10.1007/BF00171824>
- Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from pool-seq data. *Genetics*, 210(1), 315–330. <https://doi.org/10.1534/genetics.118.300900>
- Huguet, M. D., Robin, S., Hudaverdian, S., Tanguy, S., Leterme-Prunier, N., Cloteau, R., Baulande, S., Legoix-Né, P., Legeai, F., Simon, J.-C., Jaquiéry, J., Tagu, D., & Le Trionnaire, G. Transcriptomic analyses of the genetic bases of sex loss in the pea aphid [Manuscript submitted for publication].
- Janko, K., Drozd, P., Flegr, J., & Pannell, J. R. (2008). Clonal turnover versus clonal decay: a null model for observed patterns of asexual longevity, diversity and distribution. *Evolution*, 62(5), 1264–1270. <https://doi.org/10.1111/j.1558-5646.2008.00359.x>
- Jaquiéry, J., Stoeckel, S., Larose, C., Nouhaud, P., Risper, C., Mieuze, L., Bonhomme, J., Mahéo, F., Legeai, F., Gauthier, J.-P., Prunier-Leterme, N., Tagu, D., & Simon, J.-C. (2014). Genetic control of contagious asexuality in the pea aphid. *PLOS Genetics*, 10(12), 12. <https://doi.org/10.1371/journal.pgen.1004838>
- Jaquiéry, J., Stoeckel, S., Risper, C., Mieuze, L., Legeai, F., & Simon, J.-C. (2012). Accelerated evolution of sex chromosomes in aphids, an XO system. *Molecular Biology and Evolution*, 29(2), 837–847. <https://doi.org/10.1093/molbev/msr252>

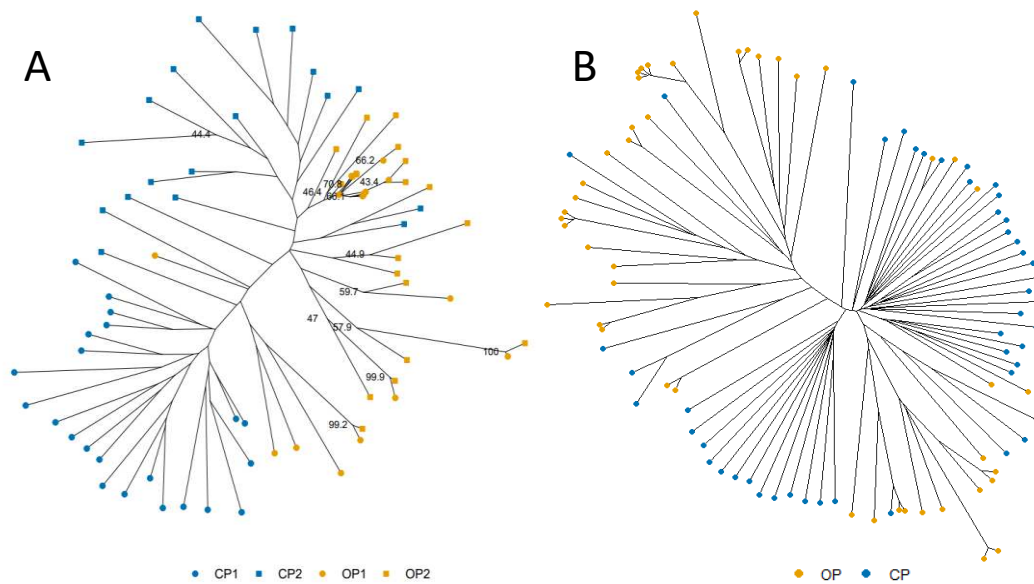
- Jaron, K. S., Parker, D. J., Anselmetti, Y., Tran Van, P., Bast, J., Dumas, Z., Figuet, E., François, C. M., Hayward, K., Rossier, V., Simion, P., Robinson-Rechavi, M., Galtier, N., & Schwander, T. (2022). Convergent consequences of parthenogenesis on stick insect genomes. *Science Advances*, 8(8), eabg3842. <https://doi.org/10.1126/sciadv.abg3842>
- Johnson, S. G., & Bragg, E. (1999). Age and polyphyletic origins of hybrid and spontaneous parthenogenetic *Campeloma* (gastropoda: Viviparidae) from the Southeastern United States. *Evolution*, 53(6), 1769–1781. <https://doi.org/10.1111/j.1558-5646.1999.tb04561.x>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>
- Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, 27(24), 3435–3436. <https://doi.org/10.1093/bioinformatics/btr589>
- Kondrashov, A. S. (1993). Classification of hypotheses on the advantage of amphimixis. *Journal of Heredity*, 84(5), 372–387. <https://doi.org/10.1093/oxfordjournals.jhered.a111358>
- Le Trionnaire, G., Francis, F., Jaubert-Possamai, S., Bonhomme, J., De Pauw, E., Gauthier, J.-P., Haubruge, E., Legeai, F., Prunier-Leterme, N., Simon, J.-C., Tanguy, S., & Tagu, D. (2009). Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid. *BMC Genomics*, 10(1), 456. <https://doi.org/10.1186/1471-2164-10-456>
- Le Trionnaire, G., Hardie, J., Jaubert-Possamai, S., Simon, J.-C., & Tagu, D. (2008). Shifting from clonal to sexual reproduction in aphids: Physiological and developmental aspects. *Biology of the Cell*, 100(8), 441–451. <https://doi.org/10.1042/BC20070135>
- Le Trionnaire, G., Jaubert-Possamai, S., Bonhomme, J., Gauthier, J.-P., Guernec, G., Le Cam, A., Legeai, F., Monfort, J., & Tagu, D. (2012). Transcriptomic profiling of the reproductive mode switch in the pea aphid in response to natural autumnal photoperiod. *Journal of Insect Physiology*, 58(12), 1517–1524. <https://doi.org/10.1016/j.jinsphys.2012.07.009>
- Lehto, M. P., & Haag, C. R. (2010). Ecological differentiation between coexisting sexual and asexual strains of *Daphnia pulex*. *Journal of Animal Ecology*, 79(6), 1241–1250. <https://doi.org/10.1111/j.1365-2656.2010.01726.x>
- Lehtonen, J., Jennions, M. D., & Kokko, H. (2012). The many costs of sex. *Trends in Ecology & Evolution*, 27(3), 172–178. <https://doi.org/10.1016/j.tree.2011.09.016>
- Li, B., Bickel, R. D., Parker, B. J., Saleh Ziabari, O., Liu, F., Vellichirammal, N. N., Simon, J.-C., Stern, D. L., & Brisson, J. A. (2020). A large genomic insertion containing a duplicated follistatin gene is linked to the pea aphid male wing dimorphism. *eLife*, 9, e50608. <https://doi.org/10.7554/eLife.50608>
- Lynch, M., & Gabriel, W. (1983). Phenotypic evolution and parthenogenesis. *The American Naturalist*, 122(6), 745–764. <https://doi.org/10.1086/284169>
- Margaritopoulos, J. T., Malarky, G., Tsitsipis, J. A., & Blackman, R. L. (2007). Microsatellite DNA and behavioural studies provide evidence of host-mediated speciation in *Myzus persicae* (Hemiptera: Aphididae). *Biological Journal of the Linnean Society*, 91(4), 687–702. <https://doi.org/10.1111/j.1095-8312.2007.00828.x>
- Mathers, T. C., Wouters, R. H. M., 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>
- Maynard Smith, J. (1971). *The origin and maintenance of sex* (Group selection). Williams GC (Ed.).
- Miyazaki, M. (1987). Forms and morphs of aphids. In A. K. Minks & P. Harrewijn (Eds.), *Aphids, their biology, natural enemies and control* (pp. 163–195).
- Moran, N. A. (1992). The evolution of aphid life cycles. *Annual Review of Entomology*, No. 37, 321–348.

- Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1), 2–9. [https://doi.org/10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8)
- Neiman, M., & Linksvayer, T. A. (2006). The conversion of variance and the evolutionary potential of restricted recombination. *Heredity*, 96(2), 2. <https://doi.org/10.1038/sj.hdy.6800772>
- Nouhaud, P., Peccoud, J., Mahéo, F., Mieuze, L., Jaquiéry, J., & Simon, J.-C. (2014). Genomic regions repeatedly involved in divergence among plant-specialized pea aphid biotypes. *Journal of Evolutionary Biology*, 27(9), 2013–2020. <https://doi.org/10.1111/jeb.12441>
- Olazuaga, L., Loiseau, A., Parrinello, H., Paris, M., Fraimout, A., Guedot, C., Diepenbrock, L. M., Kenis, M., Zhang, J., Chen, X., Borowiec, N., Facon, B., Vogt, H., Price, D. K., Vogel, H., Prud'homme, B., Estoup, A., & Gautier, M. (2020). A whole-genome scan for association with invasion success in the fruit fly *Drosophila suzukii* using contrasts of allele frequencies corrected for population structure. *Molecular Biology and Evolution*, 37(8), 2369–2385. <https://doi.org/10.1093/molbev/msaa098>
- Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Parker, D. J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M., & Schwander, T. (2019). Repeated evolution of asexuality involves convergent gene expression changes. *Molecular Biology and Evolution*, 36(2), 350–364. <https://doi.org/10.1093/molbev/msy217>
- Peccoud, J., Figueroa, C. C., Silva, A. X., Ramirez, C. C., 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(21), 4608–4618. <https://doi.org/10.1111/j.1365-294X.2008.03949.x>
- Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences*, 106(18), 7495–7500. <https://doi.org/10.1073/pnas.0811117106>
- Pons, I., Scieur, N., Dhondt, L., Renard, M.-E., Renoz, F., & Hance, T. (2022). Pervasiveness of the symbiont *Serratia symbiotica* in the aphid natural environment: Distribution, diversity and evolution at a multitrophic level. *FEMS Microbiology Ecology*, 98(1), fiac012. <https://doi.org/10.1093/femsec/fiac012>
- Prevosti, A., Ocaña, J., & Alonso, G. (1975). Distances between populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. *Theoretical and Applied Genetics*, 45(6), 231–241. <https://doi.org/10.1007/BF00831894>
- Rimbault, M., Legeai, F., Peccoud, J., Mieuze, L., Call, E., Nouhaud, P., Defendini, H., Mahéo, F., Marande, W., Théron, N., Tagu, D., Le Trionnaire, G., Simon, J.-C., & Jaquiéry, J. (2023). Contrasting evolutionary patterns between sexual and asexual lineages in a genomic region linked to reproductive mode variation in the pea aphid. *Genome Biology and Evolution*, 15(9), evad168. <https://doi.org/10.1093/gbe/evad168>
- Rispe, C., Pierre, J.-S., Simon, J.-C., & Gouyon, P.-H. (1998). Models of sexual and asexual coexistence in aphids based on constraints. *Journal of Evolutionary Biology*, 11(6), 685–701. <https://doi.org/10.1046/j.1420-9101.1998.11060685.x>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 1. <https://doi.org/10.1038/nbt.1754>
- Roy, L., Barrès, B., Capderrey, C., Mahéo, F., Micoud, A., Hullé, M., & Simon, J.-C. (2022). Host plants and insecticides shape the evolution of genetic and clonal diversity in a major aphid crop pest. *Evolutionary Applications*, 15(10), 1653–1669. <https://doi.org/10.1111/eva.13417>
- Roy, L., Fontaine, S., Caddoux, L., Micoud, A., & Simon, J.-C. (2013). Dramatic changes in the genotypic frequencies of target insecticide resistance in French populations of *Myzus persicae* (Hemiptera:

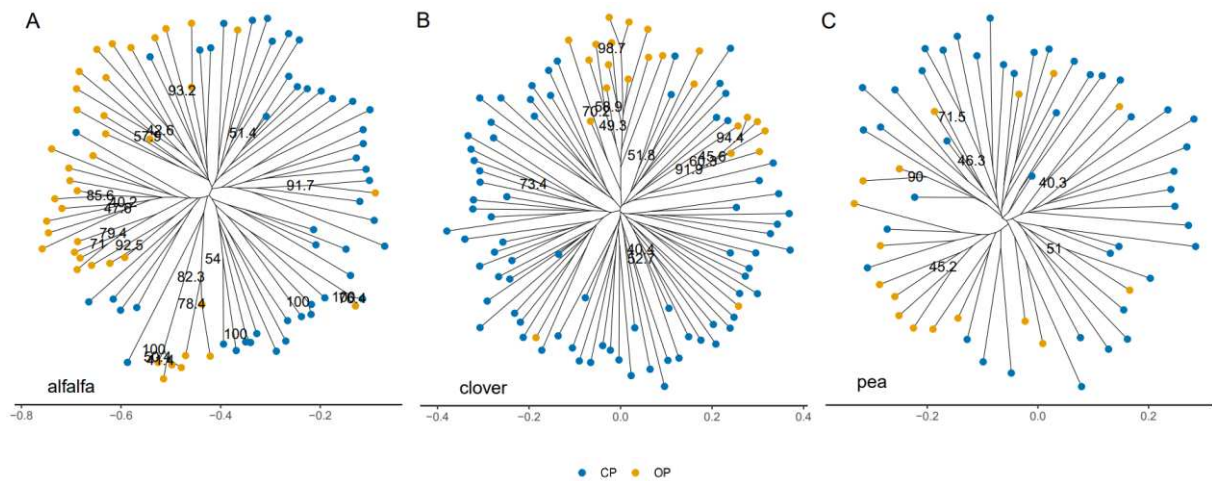
- Aphididae*) over the last decade. *Journal of Economic Entomology*, 106(4), 1838–1847. <https://doi.org/10.1603/EC12475>
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., & Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 7164. <https://doi.org/10.1038/nature06250>
- Sandrock, C., & Vorburger, C. (2011). Single-locus recessive inheritance of asexual reproduction in a parasitoid wasp. *Current Biology*, 21(5), 433–437. <https://doi.org/10.1016/j.cub.2011.01.070>
- Sayols, S. (2023). rrvgo: A Bioconductor package for interpreting lists of Gene Ontology terms. *MicroPublication Biology*, 10.17912/micropub.biology.000811. <https://doi.org/10.17912/micropub.biology.000811>
- Schön, I., Martens, K., & Dijk, P. (Eds.). (2009). *Lost Sex: The Evolutionary Biology of Parthenogenesis*. Springer Netherlands. <https://doi.org/10.1007/978-90-481-2770-2>
- Schwander, T., & Crespi, B. J. (2009). Multiple direct transitions from sexual reproduction to apomictic parthenogenesis in *Timema* stick insects. *Evolution*, 63(1), 84–103. <https://doi.org/10.1111/j.1558-5646.2008.00524.x>
- Schwander, T., Henry, L., & Crespi, B. J. (2011). Molecular evidence for ancient asexuality in *Timema* stick insects. *Current Biology*, 21(13), 1129–1134. <https://doi.org/10.1016/j.cub.2011.05.026>
- Ségurel, L., Thompson, E. E., Flutre, T., Lovstad, J., Venkat, A., Margulis, S. W., Moysse, J., Ross, S., Gamble, K., Sella, G., Ober, C., & Przeworski, M. (2012). The ABO blood group is a trans-species polymorphism in primates. *Proceedings of the National Academy of Sciences*, 109(45), 18493–18498. <https://doi.org/10.1073/pnas.1210603109>
- Simon, J. C., Blackman, R. L., & Gallic, J. F. L. (1991). Local variability in the life cycle of the bird cherry-oat aphid, *Rhopalosiphum padi* (Homoptera: Aphididae) in western France. *Bulletin of Entomological Research*, 81(3), 315–322. <https://doi.org/10.1017/S0007485300033599>
- Simon, J.-C., Baumann, S., Sunnucks, P., Hebert, P. D. N., Pierre, J.-S., Le Gallic, J.-F., & Dedryver, C.-A. (1999). Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, 8(4), 531–545. <https://doi.org/10.1046/j.1365-294x.1999.00583.x>
- Simon, J.-C., Boutin, S., Tsuchida, T., Koga, R., Gallic, J.-F. L., Frantz, A., Outreman, Y., & Fukatsu, T. (2011). Facultative Symbiont Infections Affect Aphid Reproduction. *PLOS ONE*, 6(7), e21831. <https://doi.org/10.1371/journal.pone.0021831>
- Simon, J.-C., Carrel, E., Hebert, P. D. N., Dedryver, C. A., Bonhomme, J., & Gallic, J.-F. L. (1996). Genetic diversity and mode of reproduction in French populations of the aphid *Rhopalosiphum padi* L. *Heredity*, 76(3), 3. <https://doi.org/10.1038/hdy.1996.44>
- Simon, J.-C., Delmotte, F., Rispe, C., & Crease, T. (2003). Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society*, 79(1), 151–163. <https://doi.org/10.1046/j.1095-8312.2003.00175.x>
- Simon, J.-C., Rispe, C., & Sunnucks, P. (2002). Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, 17(1), 1. [https://doi.org/10.1016/S0169-5347\(01\)02331-X](https://doi.org/10.1016/S0169-5347(01)02331-X)
- Simon, J.-C., Sakurai, M., Bonhomme, J., Tsuchida, T., Koga, R., & Fukatsu, T. (2007). Elimination of a specialised facultative symbiont does not affect the reproductive mode of its aphid host. *Ecological Entomology*, 32(3), 296–301. <https://doi.org/10.1111/j.1365-2311.2007.00868.x>
- Simon, J.-C., Stoeckel, S., & Tagu, D. (2010). Evolutionary and functional insights into reproductive strategies of aphids. *Comptes Rendus Biologies*, 333(6), 6. <https://doi.org/10.1016/j.crv.2010.03.003>

- Singh, K. S., Cordeiro, E. M. G., Troczka, B. J., Pym, A., Mackisack, J., Mathers, T. C., Duarte, A., Legeai, F., Robin, S., Bielza, P., Burrack, H. J., Charaabi, K., Denholm, I., Figueroa, C. C., French-Constant, R. H., Jander, G., Margaritopoulos, J. T., Mazzoni, E., Nauen, R., ... 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), 1. <https://doi.org/10.1038/s42003-021-02373-x>
- Sperling, A. L., Fabian, D. K., Garrison, E., & Glover, D. M. (2023). A genetic basis for facultative parthenogenesis in *Drosophila*. *Current Biology*, *33*(17), 3545-3560.e13. <https://doi.org/10.1016/j.cub.2023.07.006>
- Stelzer, C.-P., Schmidt, J., Wiedlroither, A., & Riss, S. (2010). Loss of sexual reproduction and dwarfing in a small Metazoan. *PLOS ONE*, *5*(9), e12854. <https://doi.org/10.1371/journal.pone.0012854>
- Szpiech, Z. A., & Hernandez, R. D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, *31*(10), 2824–2827. <https://doi.org/10.1093/molbev/msu211>
- Tang, C. Q., Obertegger, U., Fontaneto, D., & Barraclough, T. G. (2014). Sexual species are separated by larger genetic gaps than asexual species in Rotifers. *Evolution; International Journal of Organic Evolution*, *68*(10), 2901–2916. <https://doi.org/10.1111/evo.12483>
- Tougeron, K., & Iltis, C. (2022). Impact of heat stress on the fitness outcomes of symbiotic infection in aphids: A meta-analysis. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1971), 20212660. <https://doi.org/10.1098/rspb.2021.2660>
- Tucker, A. E., Ackerman, M. S., Eads, B. D., Xu, S., & Lynch, M. (2013). Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. *Proceedings of the National Academy of Sciences*, *110*(39), 15740–15745. <https://doi.org/10.1073/pnas.1313388110>
- 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*(12), 3493–3504. <https://doi.org/10.1046/j.1365-294X.2003.01998.x>
- Vrijenhoek, R. C., Dawley, R. M., Cole, C. J., & Bogart, J. P. (1989). A list of known unisexual vertebrates. In R. M. Dawley & J. P. Bogart (Eds.), *Evolution and ecology of unisexual vertebrates* (pp. 19–23). University of the State of New York.
- Vrijenhoek, R. C., & Parker, E. D. (2009). Geographical parthenogenesis: general purpose genotypes and frozen niche variation. In I. Schön, K. Martens, & P. Dijk (Eds.), *Lost Sex: The Evolutionary Biology of Parthenogenesis* (pp. 99–131). Springer Netherlands. https://doi.org/10.1007/978-90-481-2770-2_6
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), e49. <https://doi.org/10.1093/nar/gkr1293>
- Wright, S. (1977). *Evolution and the Genetics of Populations: Vol. 3 Experimental Results and Evolutionary Deductions*. University of Chicago Press.
- Xu, S., Spitze, K., Ackerman, M. S., Ye, Z., Bright, L., Keith, N., Jackson, C. E., Shaw, J. R., & Lynch, M. (2015). Hybridization and the origin of contagious asexuality in *Daphnia pulex*. *Molecular Biology and Evolution*, *32*(12), 3215–3225. <https://doi.org/10.1093/molbev/msv190>
- Yagound, B., Dogantzis, K. A., Zayed, A., Lim, J., Broekhuysse, P., Remnant, E. J., Beekman, M., Allsopp, M. H., Aamidor, S. E., Dim, O., Buchmann, G., & Oldroyd, B. P. (2020). A single gene causes thelytokous parthenogenesis, the defining feature of the Cape Honeybee *Apis mellifera capensis*. *Current Biology*, *30*(12), 2248-2259.e6. <https://doi.org/10.1016/j.cub.2020.04.033>

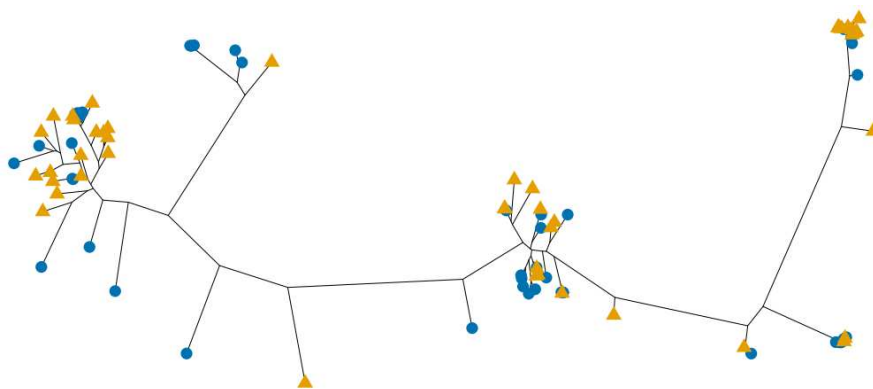
Supplementary materials for chapter 2



Suppl. Figure 1. Genetic relationships between presumably obligately parthenogenetic (OP) and cyclically parthenogenetic (CP) lineages of *Myzus persicae*. **(A)** Neighbor-joining tree based on Prevosti genetic distance between 77 genotypes over 14 microsatellite loci on the “sympatric dataset”. The numbers at the nodes are bootstrap percentages (1000 replicates). Bootstrap values under 40% are not shown. CP lineages are shown in blue and OP lineages in orange. The shape of the dots differentiates the pools (circles for CP1 and OP1, squares for CP2 and OP2). **(B)** Genetic relationships between the presumably 46 OP and 50 CP *M. persicae* lineages from the “worldwide dataset”. The neighbor-joining tree based on Prevosti genetic distance was computed using 574,677 SNPs.



Suppl. Figure 2. Genetic relationships between obligately parthenogenetic (OP) and cyclically parthenogenetic (CP) lineages of the alfalfa, pea and clover races of *Acyrthosiphon pisum*. The different neighbor-joining tree were constructed using the Prevosti genetic distance measured between **(A)** 48 CP and 42 OP lineages of alfalfa race at 14 microsatellite loci, **(B)** 45 CP and 23 OP lineages of clover race at 15 microsatellite loci, and **(C)** 43 CP and 17 OP lineages of pea race at 15 microsatellite loci. Numbers at nodes are bootstrap percentages (1000 replicates). Bootstrap values under 40% are not shown. CP lineages are shown in blue and OP lineages in orange.



Suppl. Figure 3. Genetic relationships between obligately parthenogenetic (OP) and cyclically parthenogenetic (CP) lineages of *Myzus persicae* across a candidate genomic region for sex loss. The neighbor-joining tree was based on genetic distance between the 96 individuals across a region on chromosome A3, containing 350 SNPs, and potentially associated with sex loss (peak 1 in Figure 3). The 46 OP lineages are in orange and the 50 CP lineages are in blue.



Supplementary tables are available on

<https://drive.google.com/drive/folders/1ujOkzfiNIU3EWqthAqciDCXRdrD1Y-R?usp=sharing>

Chapter 3

The third chapter of my PhD deals with the evolutionary consequences of the transitions from cyclical to obligate asexuality on sexual traits in the pea aphid. I tested for evidence of relaxation of selection on traits related to male production and male fitness in obligately parthenogenetic lineages of the pea aphid. I started this study during my Master 2 internship (carried out in the same team as my PhD thesis). However due to the COVID-19 lockdown, some of the data (notably all the wet-lab genetic analyses) was produced during my PhD. This manuscript has been published in *Molecular Ecology*.

Evolutionary consequences of loss of sexual reproduction on male-related traits in parthenogenetic lineages of the pea aphid

Hélène Defendini¹  | Maud Rimbault¹ | Frédérique Mahéo¹ | Romuald Cloteau¹ | Gaëtan Denis¹ | Lucie Mieuzet¹ | Yannick Outreman² | Jean-Christophe Simon¹ | Julie Jaquiéry¹ 

¹UMR 1349 IGEPP, INRAE, Institut Agro, Université Rennes 1, Le Rheu, France

²UMR 1349 IGEPP, INRAE, Institut Agro, Université Rennes 1, Rennes, France

Correspondence

Julie Jaquiéry, UMR 1349 IGEPP, INRAE, 35653 Le Rheu, France.

Email: julie.jaquier@inrae.fr

Funding information

French Research Agency (ANR), Grant/Award Number: ANR-09-GENM-017-001 and ANR-11-BSV7-005-01; INRAE-SPE Department; Région Bretagne

Handling Editor: Tatiana Giraud

Abstract

Transition from sexual reproduction to parthenogenesis constitutes a major life-history change with deep evolutionary consequences for sex-related traits, which are expected to decay. The pea aphid *Acyrtosiphon pisum* shows intraspecific reproductive polymorphism, with cold-resistant cyclically parthenogenetic (CP) lineages that alternate sexual and asexual generations and cold-sensitive obligately parthenogenetic (OP) lineages that produce only asexual females but still males. Here, the genotyping of 219 pea aphid lineages collected in cold-winter and mild-winter regions revealed contrasting population structures. Samples from cold-winter regions consisted mostly of distinct multilocus genotypes (MLGs) usually represented by a single sample (101 different MLGs for 111 samples) and were all phenotyped as CP. In contrast, fewer MLGs were found in mild-winter regions (28 MLGs for 108 samples), all but one being OP. Since the males produced by OP lineages are unlikely to pass on their genes (sexual females being rare in mild-winter regions), we tested the hypothesis that their traits could degenerate due to lack of selection by comparing male production and male reproductive success between OP and CP lineages. Male production was indeed reduced in OP lineages, but a less clear pattern was observed for male reproductive success: females mated with OP males laid fewer eggs (fertilized or not) but OP and CP males fertilized the same proportion of eggs. These differences may stem from the type of selective forces: male production may be counter-selected whereas male performances may evolve under the slower process of relaxed selection. The overall effective reproductive capacity of OP males could result from recent sex loss in OP lineages or underestimated reproductive opportunities.

KEYWORDS

clonal evolution, loss of sexuality, male sexual traits, relaxed selection, reproductive success, trait decay

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Sexual reproduction is thought to have evolved in the common ancestor of eukaryotes, as it is highly frequent in this domain of life (Goodenough & Heitman, 2014). However, sexual reproduction was secondarily and independently lost in many groups, with loss of sex being occasional in protists, plants and fungi and rare in vertebrates (Bell, 1982; Simon et al., 2003; Vrijenhoek et al., 1989).

The transition from obligate sexual reproduction to parthenogenesis has many evolutionary consequences for phenotypic and genotypic characteristics. In particular, traits that evolved under sexual reproduction are expected to decay through selection or drift, depending on the cost of expressing the trait. As such, males produced by parthenogenetic females represent an energetic cost without any associated benefit when they are isolated from sexually reproducing populations. We thus expect natural selection to favour parthenogenetic females allocating fewer resources to male production. This has indeed been documented in several parthenogenetic species, including snails (Neiman et al., 2012) and *Daphnia* (Innes et al., 2000). Similarly, allocation into the male function in hermaphrodite mussel species is lower in populations with less outcrossing opportunities (Johnston et al., 1998).

When males still occur in exclusively parthenogenetic populations isolated from sexual counterparts, they are no longer under selection. No matter the potential reproductive performance of the males, they will not pass on their genes. Therefore, in the absence of pleiotropy, deleterious mutations are expected to accumulate through the slow process of drift leading to male sexual trait decay (Kraaijeveld et al., 2016). Evidence of degeneration is found in sperm production and morphology in several parthenogenetic species, including the snail *Potamopyrgus antipodarum* (Jalinsky et al., 2020), the parasitoid wasp *Muscidifurax uniraptor* (Gottlieb & Zchori-Fein, 2001) and the springtail *Folsomia candida* (Kampfraath et al., 2020). The ability to fertilize sexual females is altered at different degrees in males of several thelytokous parasitoid wasps (Adachi-Hagimori & Miura, 2020; Pannebakker et al., 2005; Zchori-Fein et al., 1992) as well as in asexual stick insects (Schwander et al., 2013).

Despite this evidence of degeneration, males produced by parthenogenetic species remain globally fertile (van der Kooi & Schwander, 2014). This may be explained by a recent loss of sex in these lineages, leaving insufficient time for relaxed selection to operate. Furthermore, the power of these hypotheses as explanations for the decay of male traits depends on the extent to which sons of parthenogenetic mothers have mating opportunities. Even a low rate of sex may be sufficient to prevent mutational degeneration of male genes as demonstrated by modelling in a hermaphrodite nematode (Chasnov & Chow, 2002), and rare opportunities for reproduction with sexual females may be particularly difficult to detect.

Aphids are appropriate organisms to test for sexual trait decay following sex loss as these insects present various modes of reproduction. The ancestral reproductive mode of this group, which evolved around 280 million years ago, is cyclical parthenogenesis (Grimaldi & Engel, 2005). This peculiar life-cycle consists of

an alternation between many generations of parthenogenesis by viviparous parthenogenetic females in spring and summer and a sexual generation in autumn with males and oviparous sexual females (Dixon, 1998; Hales et al., 1997). The production of the sexual morphs is triggered by the autumnal decrease in daylight length (Simon et al., 2010). An XX/X0 sex determining system allows the production of males by parthenogenesis, with the elimination of one of the X chromosomes to produce a male (Blackman, 1975). Then, as males produce only X-bearing sperms, the fusion of male and female gametes leads to a diploid individual at the X, which develops into a parthenogenetic female.

Interestingly, several aphid species—including the pea aphid *Acyrtosiphon pisum* (Harris, 1776)—show genetically determined reproductive polymorphism, with lineages that are cyclically parthenogenetic (CP lineages) and others that have lost the ability to reproduce sexually (Dedryver et al., 2013; Jaquiéry et al., 2014; Simon et al., 2010). These obligately parthenogenetic lineages (OP) are not able to produce sexual females though they often retain the ability to produce males in response to day length shortening. Only CP lineages can survive cold winters because they produce oviparous sexual females, which lay frost-resistant eggs and remain in diapause for three to four months (Simon et al., 2002). In contrast, OP lineages are cold-sensitive and prevail in warmer areas, owing to their ability to maintain parthenogenetic development even in winter, which gives them a demographic advantage over CP lineages in such conditions (Simon et al., 2002). This divergent selection on the mode of reproduction by climate thus generally induces a geographical separation of the sexual (CP) and asexual (OP) lineages (Dedryver et al., 2001; Papura et al., 2003; Simon et al., 1999; Vorburger et al., 2003). Local adaptation to climate probably facilitated the long-term coexistence of these two reproductive modes (Simon et al., 2010).

Gene flow between OP and CP lineages is nevertheless likely to occur in different aphid species. First, the two types of lineages are expected to be sympatric in regions with intermediate climates (Dedryver et al., 2001; Rispe et al., 1998; Rispe & Pierre, 1998). The high dispersal ability of aphids could further favour the co-occurrence of OP and CP lineages even in less suitable habitats. Indeed, parthenogenetic females can develop wings under certain environmental conditions, and thus may be passively wind-dispersed (Loxdale et al., 1993). Second, although it is difficult to demonstrate by direct observation that OP males are also produced in nature (but see Halkett et al., 2008), we know that OP aphid lineages can produce males in laboratory conditions in response to sex-inducing cues. OP males from the laboratory have been successfully mated with sexual females (Blackman, 1972; Dedryver et al., 2012; Jaquiéry et al., 2014), suggesting that crossings in the wild may also be achieved. Indirect evidence of gene flow between OP and CP lineages is supported by population genetic data: in the alfalfa host race of the pea aphid complex, OP and CP populations more than 450 km apart show low genome-wide genetic differentiation ($F_{ST}=2.4\%$), except at an 840 kb genomic region associated with the reproductive phenotype (Jaquiéry et al., 2014; Rimbault et al., 2022). This suggests that gene flow between OP and CP lineages almost

counterbalances genetic drift—which is probably moderate given the large sizes of aphid populations living on legume crops—and thus prevents genetic differentiation in genomic regions not linked with reproductive polymorphism.

Here, we first aimed at characterizing the genotypic diversity and genetic structure of clover-adapted pea aphid populations sampled in two regions of France with contrasting winter climates. By characterizing the reproductive phenotype of nearly a 100 lineages, we demonstrated extensive geographic segregation of the two reproductive modes, suggesting that males from OP lineages—which do not produce sexual females—may rarely encounter sexual females. In such a situation, male-related traits could be expected to degenerate as a result of relaxed selection or counter-selection. Our second aim was thus to investigate the fate of male traits—including male production and male mating success—in OP compared to CP lineages. The reproductive success of OP and CP males was measured with and without male-male competition to explore a range of conditions that might accentuate potential differences in performance. Finally, if OP lineages represent independent transitions to obligate parthenogenesis, their respective time since sex loss may differ, which could affect the intensity of degeneration of male traits. We thus examined the genetic relatedness between the OP and CP lineages to assess whether OP lineages derived from single or multiple sex loss events were closely related to CP lineages, which would suggest recent sex loss. We then tried to relate the genetic relationships between CP and OP lineages and the male trait performances of the different OP lineages.

2 | MATERIALS AND METHODS

2.1 | Aphid lineages and genetic analyses

A total of 563 *A. pisum* individuals were collected in clover fields located in the east (265 individuals) or west (298 individuals) of France in 2017 or 2018. Collection sites in the eastern and western regions were at least 480 km apart, while sites within the same region were 26.6 km apart on average (Figure 1, Table S1). These two regions were chosen for their contrasting winter temperatures, with mild winters in the west due to the nearby ocean and colder winters in the east. Each collected aphid was installed individually on a faba bean (*Vicia faba*) plant kept at 18°C with a 16:8 light:dark cycle to maintain parthenogenetic reproduction. Only 219 of the 563 individuals successfully established a clonal colony in the laboratory because more than 50% of wild caught individuals were parasitized by entomopathogenic fungi or parasitoid wasps and died without producing offspring. These 219 lineages (108 from the west, 111 from the east) were then genotyped at 15 polymorphic microsatellite loci (AIA09M, AIB07M, AIB08M, AIB12M, ApF08M, ApH08M, ApH10M, X111775, X114782, X117294, X118775, X118777, X126104, X127526, X128001, Caillaud et al., 2004, Jaquiéry et al., 2012, Nouhaud et al., 2014) following DNA extraction and PCR protocols described in Peccoud et al. (2008). Lineages that had

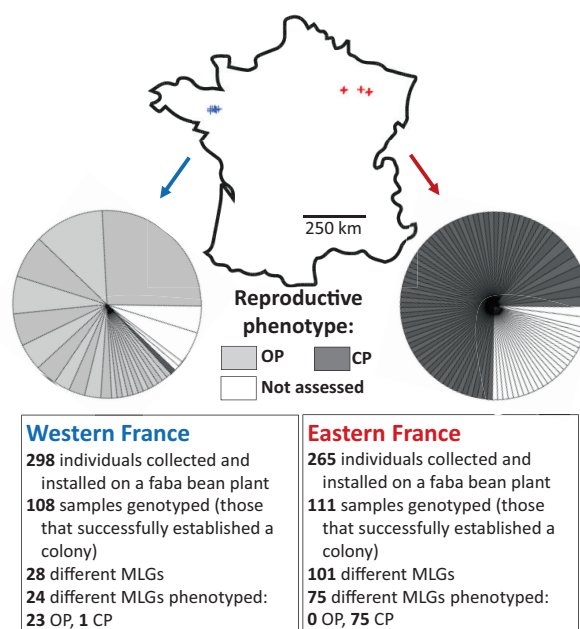


FIGURE 1 Genotypic diversity of pea aphid (*Acyrtosiphon pisum*) samples collected in western (blue crosses) and eastern France (red crosses) in clover fields. Each slice of the pie charts represents one distinct MLG (multilocus genotype at 15 microsatellite markers), its size being proportional to the number of individuals sampled in the wild (and genotyped) having this MLG. The reproductive mode of each MLG is shown in light grey for OP (obligate parthenogenesis) and dark grey for CP (cyclical parthenogenesis).

the same genotype at all 15 loci or differed by at most one allele were identified with the R package RClone (Baillieu et al., 2016) and were assumed to be clonal copies of the same multilocus genotype (MLG). Only 28 different MLGs were found in the west (out of the 108 genotyped samples), while there were 101 different MLGs in the east (out of 111 genotyped samples). This dataset is available in Table S1. We kept alive only one representative for each MLG for all subsequent analyses. A subsample of 75 lineages (out of the 101 different MLGs) from the east and 24 lineages from the west (4 lineages died prior to this point) were then characterized in the laboratory for their mode of reproduction (OP versus CP, see below for detailed protocols). Finally, we assessed the genetic differentiation (F_{ST}) between eastern and western samples (keeping only one copy of each MLG, that is, 24 from the west and 75 from the east) on the 15 microsatellite markers using the *hierfstat* package (Goudet & Jombart, 2022) in R (version 4.0.3, R Core Team, 2019). Expected heterozygosity (H_e) within each geographical region (using only one copy of each MLG) was computed using the *adegenet* package (Jombart, 2008) and the difference between regions tested with the function *Hs.test* (with 999 permutations). A principal component analysis (PCA) was also performed on lineages using the *ade4* R package (Bougeard & Dray, 2018). Finally, we assessed the genetic relationship between OP and CP lineages using the shared allele distance D_{AS} (Chakraborty & Jin, 1993). A neighbour-joining tree

based on D_{AS} was built using the *ggtree* R package (Yu et al., 2017), and bootstrap values were computed over 1000 replications with the *aboot* function (Kamvar et al., 2015) from the R package *poppr* (Kamvar et al., 2014).

2.2 | Characterization of the reproductive mode

To determine the reproductive mode (OP or CP) of the 99 lineages and to measure their investment in the production of the different morphs (i.e. males, sexual females and parthenogenetic females), lineages were reared under sex-inducing conditions (as described in Le Trionnaire et al., 2009). Briefly, one one-day-old larva was deposited on a *V. faba* plant stored at 18°C with a 16:8L:D photoperiod regime for 1 week and was then moved to short-day conditions (12:12L:D regime, 18°C). Six days later, one of its offspring (i.e. the second generation) was collected, placed on a new plant and maintained under short-day conditions. Once it had reached adulthood, this second-generation individual is expected to first produce females (oviparous sexual females for CP lineages, parthenogenetic females for OP lineages), then males and finally parthenogenetic females. The second-generation individual was transferred onto a new plant every week (to avoid crowding due to the large number of offspring produced). Offspring were placed at 18°C under 16:8L:D photoperiod, and their reproductive status (sexual female, male or parthenogenetic female) was determined individually based on morphological characters (Miyazaki, 1987) once they were adults. For each of the 99 lineages, three replicates of this experiment were performed. These analyses revealed that all the 75 lineages from the east were CP, and that all but one (i.e. 23) from the west were OP (relevant data is fully available in Table S2).

To assess whether the proportion of males produced by a lineage depends on its reproductive mode, a generalized linear mixed model (GLMM) assuming a binomial error and a logit-link function was fitted using the *glmer* function from the *lme4* package (Bates et al., 2015). We considered the reproductive mode of the lineage as a two-level fixed factor and the lineage as a nested random factor within the reproductive mode (Equation 1). This random factor extracts the influence of aphid genotype on the dependent variables and removes any confounding effect between reproductive mode and aphid genotype, while accounting for variability within the reproductive mode. We added an observation-level random effect (OLRE, Harrison, 2014, 2015) to deal with overdispersion.

$$\text{logit}(\text{proportion of males}) \sim \text{reproductive mode} + (1|\text{reproductive mode: male lineage}) + (1|\text{OLRE}) \quad (1)$$

The best model was selected based on term significance, which was determined by a Wald Chi-square test as a type III analysis of variance using the *car* package *Anova* function (Fox & Weisberg, 2018). Traditional R^2 measures are not well suited to measuring the goodness of fit of mixed-effects models, which include random-effects variance components. We therefore calculated two appropriate

measures of R^2 for mixed-effects models on the final model including only significant variables: the marginal R^2 (the variance explained by the fixed effects) and the conditional R^2 (the variance explained by both fixed and random effects, Nakagawa & Schielzeth, 2013). This was done using the r^2_{nakagawa} function from the R package *performance* (Lüdtke et al., 2021).

2.3 | Mating experiments without male-male competition

Mating experiments were conducted on the 24 lineages (12 OP and 12 CP lineages) that were still maintained in the laboratory at the start of the experiment (early 2020, Tables S1 and S3). These 24 lineages had been kept alive since 2018 because they were representative of the diversity of the 99 lineages. To produce a large number of males required for mating experiments, sex-inducing conditions similar to those described above were used. However, to optimize the production of males, tightly control their age and minimize handling, four individuals of the second-generation (those producing sexual morphs) were maintained on the same plant (instead of one per plant previously). The four individuals were transferred onto a new plant every one or two days (depending on the number of larvae laid) to avoid crowding. Two CP lineages (CP4 and CP10) were used to produce sexual females for mating experiments using the same protocol. We chose to work with two female genotypes to account for a possible female effect, but we could not test more due to handling constraints. By isolating males and females prior to sexual maturation (at the fourth larval stage), we ensured that sexual morphs were virgin at the start of experiments.

We first assessed the individual reproductive success of OP and CP males when alone with five sexual female partners (1:5 male-to-female ratio). We used 13- to 14-day-old virgin wingless males and females (hence sexually mature, Sack & Stern, 2007). First, we placed five females of the same lineage (either CP4 or CP10) and one male of each of the 24 lineages (either a CP or an OP lineage) on a faba bean plant grown in a tube filled with soil (2 cm diameter, 9.5 cm long). Plants were cut at the apical bud (leaving only the first two leaves) so that they all had similar leaf area and morphology allowing more contacts between male and females. The soil was covered with parafilm, to facilitate egg recovery at the end of the experiment. The plant was covered with a small drilled bag and this experimental setup was left untouched for 7 days under long day conditions (16 h of light, 18°C). After 7 days, the six individuals were removed from the plant. Fertilized and viable eggs laid by *A. pisum* females turn black through melanization after a few days whereas unfertilized or unviable eggs stay green (Blackman, 1987). Five days after removal of the adults (a sufficient time for the eggs to melanize), the total number of eggs and the proportion of melanized eggs found on the plant or on the bag were recorded. All sexual females used in the mating experiments were virgins and of the same age, and the abdomen of each was already full with many mature eggs and other developing eggs that were visible through the cuticle. Sexual pea

aphid females' ovaries contain 14 ovarioles (Wieczorek et al., 2019) and one ovariole contains 32 germ cells, 11 of them developing into oocytes (Blackman, 1987; Miura et al., 2003). Sexual females can therefore lay a large number of eggs throughout their adult life, indicating that the number of eggs per female is not a limiting factor in our study. We observed that sexual females laid more eggs when they were in contact with more males (see results), which suggests that males (or mating) stimulate egg laying. Therefore, the total number of eggs laid by females throughout 7 days with males may indicate male's ability to stimulate females, whereas the proportion of melanized eggs may indicate the quality of the sperm. We performed 246 replicates of the mating experiment on the 24 lineages, with 1 to 27 replicates per lineage depending upon the availability of males. A total of 144 males from the 12 CP lineages and 102 males from the 12 OP lineages were tested. Approximately half of the OP males and half of the CP males were tested with CP4 females and the other two halves with CP10 females. The majority of the mating experiments were carried out entirely in the laboratory, while 42.3% (44.1% for OP males, and 41.0% for CP males) were partly or entirely carried out outside the laboratory, in a room with less controlled conditions (12:12 L:D regime, ~18–20°C) due to the COVID-19 lockdown and closure of the laboratory. This factor was accounted for in our analyses. Few adult individuals died before the end of the 7-day experiment (9 males out of 242, i.e. 3.7% and 17 out of 1210 females, i.e. 1.4%). Relevant data is fully available in Table S3.

The effect of the reproductive mode of the male lineage on the response variables (i.e. the number of eggs laid and the proportion of fertilized eggs) was tested using GLMM (negative binomial error with log-link function and binomial error with logit-link function resp.), including the female lineage (two levels: CP4 or CP10) and the location of the experiment (two levels: in the laboratory or out the laboratory) as fixed factors, and the male lineage as a random factor nested within the reproductive mode (Equations 2 and 3). In these two models, we added the interaction between the reproductive mode of the male lineage and the lineage of the female, which estimated whether the effect of the reproductive mode of the male on its reproductive success may vary according to the female lineage. In the model of the proportion of fertilized eggs, we included an observation-level random effect (OLRE, Harrison, 2014, 2015) to deal with overdispersion. Four tubes with no eggs (three cases with CP males and one with an OP male) were excluded for the analysis of the proportion of fertilized eggs.

$$\log(\text{number of eggs laid}) \sim \text{reproductive mode} + \text{female lineage} + \text{location} + \text{reproductive mode} \times \text{female lineage} + (1|\text{reproductive mode} : \text{male lineage}) \quad (2)$$

$$\text{logit}(\text{proportion of fertilized eggs}) \sim \text{reproductive mode} + \text{female lineage} + \text{location} + \text{reproductive mode} \times \text{female lineage} + (1|\text{reproductive mode} : \text{male lineage}) + (1|\text{OLRE}) \quad (3)$$

The variance explained by the different types of variables was measured through marginal and conditional R^2 as previously described.

2.4 | Mating experiments under male–male competition

To test the reproductive competitiveness of CP and OP males, we compared their respective reproductive success in a situation of strong male–male competition for females (10:5 male-to-female ratio). This competition experiment involved five males from different OP lineages (lineages OP1, OP2, OP3, OP4 and OP5), five males from four different CP lineages (CP1, CP2, CP3 and two males from CP4) and five females of the CP10 lineage. We only used CP10 as the female-producing lineage in this mating experiment as it produced the highest number of sexual females and could therefore be more easily synchronized with male production. The lineages used to produce males for this experiment were chosen because a sufficient number of same-aged males (13- to 14-day-old) was simultaneously available to perform six replicates. The even number of CP and OP males leads to the same competitive pressure between OP and CP males. However, two males from the same CP4 lineages (instead of one from CP4 and one from CP9) were mistakenly included in the six replicates of this mating experiment. This was considered in the statistical analysis. The 10 males were placed altogether with the five females on a faba bean plant cut at the apical bud, covered with a drilled bag and with soil covered with parafilm. After 7 days, all aphids were removed from the plant and preserved in 96% ethanol for microsatellite genotyping, in order to confirm their genotype. Among the six replicates, only two males and two females were found dead at the end of the experiment (4.4%). The counting of the total number of eggs and melanized eggs was done 5 days later. Eggs were stored individually in Eppendorf tubes at –20°C before being manually crushed with a plastic pestle in a proteinase K lysis solution. DNA extraction was then performed as in Peccoud et al. (2008). Each egg and adult were genotyped at 7 polymorphic microsatellite loci (AIA09M, AIB07M, AIB08M, AIB12M, ApF08M, ApH08M and ApH10M, Caillaud et al., 2004) following the conditions described in Peccoud et al. (2008). Each genotype is characterized by a specific combination of alleles at these highly polymorphic microsatellite loci, allowing paternity inference (Tables S1 and S4). Six replicates of this competition experiment, with exactly the same lineages, were carried out. Four of them were conducted in the laboratory while the two remaining ones were done outside the laboratory, in a room with less controlled conditions (12:12 L:D regime, ~18–20°C) due to the COVID-19 lockdown. Since two males from CP4 genotype were mistakenly put in each replicate, the number of eggs fertilized by these males was theoretically twice the number of eggs one male of this lineage could really fertilize. Therefore, half of the number of eggs fertilized by males from CP4 lineage were assigned to this lineage in the following analysis, to give the same weight to each male lineage. Relevant data is fully available in Table S4.

The proportion of eggs fertilized by one male lineage out of all fertilized eggs in a replicate was compared between reproductive modes (Equation 4) with GLMM assuming binomial error and using logit-link function. The male lineage was included in the model as a

random factor nested within the reproductive mode (Equation 4). The replicate number was included as a random effect, the two-level factor location as fixed effect, and an observation-level random effect (OLRE, Harrison, 2014, 2015) was included to deal with overdispersion.

$$\text{logit}(\text{proportion of fertilized eggs}) \sim \text{reproductive mode} + \text{location} + (1|\text{reproductive mode: male lineage}) + (1|\text{replicate}) + (1|\text{OLRE}) \quad (4)$$

The variance explained by the different types of variables was measured through marginal and conditional R^2 as previously described.

2.5 | Comparison of reproductive success with and without competition

To examine whether the fitness of males from the different lineages was consistent regardless of competition level, we calculated the Spearman correlation between the average number of fertilized eggs per male lineage with and without competition. To avoid a possible effect of the female lineage between the two mating experiments, we only considered the replicates with the female lineage CP10 (which was used in the two types of mating experiments).

3 | RESULTS

3.1 | Genetic structure and reproductive phenotype of wild-caught aphid samples

The genotypic composition of the pea aphid populations from western and eastern France clearly differed: in the east, 101 of the 111 colonies that established in the laboratory and were genotyped at 15 microsatellite loci corresponded to different multilocus genotypes (i.e. $G/N=101/111=0.91$). In the west, the 108 colonies corresponded to 28 different multilocus genotypes only (i.e. $G/N=28/108=0.26$). There was no shared MLGs between eastern and western populations. The reproductive mode of the different MLGs was assessed for 75 MLGs from eastern France (out of 101) and 24 MLGs from western France (out of 28) and differed between the two regions: under sex-inducing conditions, all the 75 different MLGs from eastern France produced sexual females, and were thus classified as CP. In contrast, all but one of the 24 different MLGs collected in western France did not produce sexual females and were thus classified as OP. When considering MLGs, the CP phenotype was extremely rare in western fields, accounting for about 1% of the collected samples (i.e. 1 out of 108 samples, Figure 1).

The next analyses were performed on the 99 different MLGs (23 OP and 76 CP). Expected heterozygosity did not differ significantly between eastern and western populations ($H_e \text{ East}=0.761$, $H_e \text{ West}=0.753$, $p=.34$). Eastern and western populations were

significantly differentiated ($F_{ST}=0.048$, 95% CI: [0.031–0.068]). The PCA separated OP and CP lineages along the first axis, which explained 7.9% of the variance. However, two OP lineages (OP1 and OP2) grouped with the cluster formed by CP lineages (Figure 2a). D_{AS} distances (mean \pm s.d.) among lineages sharing the same reproductive mode were 0.621 ± 0.007 for OP lineages and 0.659 ± 0.001 for CP lineages. This distance reached 0.711 ± 0.002 when measured between pairs of OP and CP lineages. The tree based on D_{AS} showed that most OP lineages tended to cluster together, with some well supported groups (e.g. OP3, OP10, OP12, OP16 and OP17) showing shallow branching, whereas CP lineages were generally characterized by deeper branching. Two OP lineages (OP1 and OP2, the same as in the PCA) were closer to the CP lineages (Figure 2b) although bootstrap support was low, as expected for individual genotype data based on 15 microsatellite loci.

3.2 | Male production in OP and CP lineages

We then analysed the investment of the 23 OP and 76 CP lineages into the production of males, sexual and parthenogenetic females in more detail. All three types of morph could be produced by CP lineages, whereas OP lineages never produced sexual females and produced less males and more parthenogenetic females than CP lineages (Table 1, Figures S1). The total fecundity (i.e. the sum of all morphs produced) was slightly higher for CP compared to OP lineages (GLMM, $p=.03$, Table 1). The frequency of lineages that produced males was higher in CP lineages than in OP lineages, but this was not significant (chi² test, $\chi^2=1.369$, $df=1$, $p=.24$, Table 1). Furthermore, the averaged proportion of males produced by OP lineage was smaller than for CP lineages (GLMM, $p=.015$, Tables 1 and 2, Figure 3a). The contribution of both fixed and random effects to the variance of this GLMM model was 39.9% (given by conditional R^2). This value partitioned into 3.6% due to fixed effects (i.e. the reproductive mode, given by marginal R^2) and 36.3% to random effects (i.e. the male lineage and observation-level random effects, obtained by subtracting the marginal R^2 from the conditional R^2).

3.3 | Male reproductive success without competition

The individual reproductive success of males from 12 OP and 12 CP lineages was then assessed. After seven days of contact between the male and the five females, eggs were found in 98.4% of the 246 replicates. In total, 6056 eggs were identified, of which 77% were fertilized. Fertilized eggs were found in half of the replicates involving OP lineages and two-third of those involving CP lineages (Table 3, Table S3).

Among the 246 replicates, from zero to 103 eggs were found per tube. The best-fitting GLMM explaining the number of eggs laid revealed a significant effect of the reproductive mode of the male lineage ($p=.005$), with fewer eggs laid by females sired by OP males

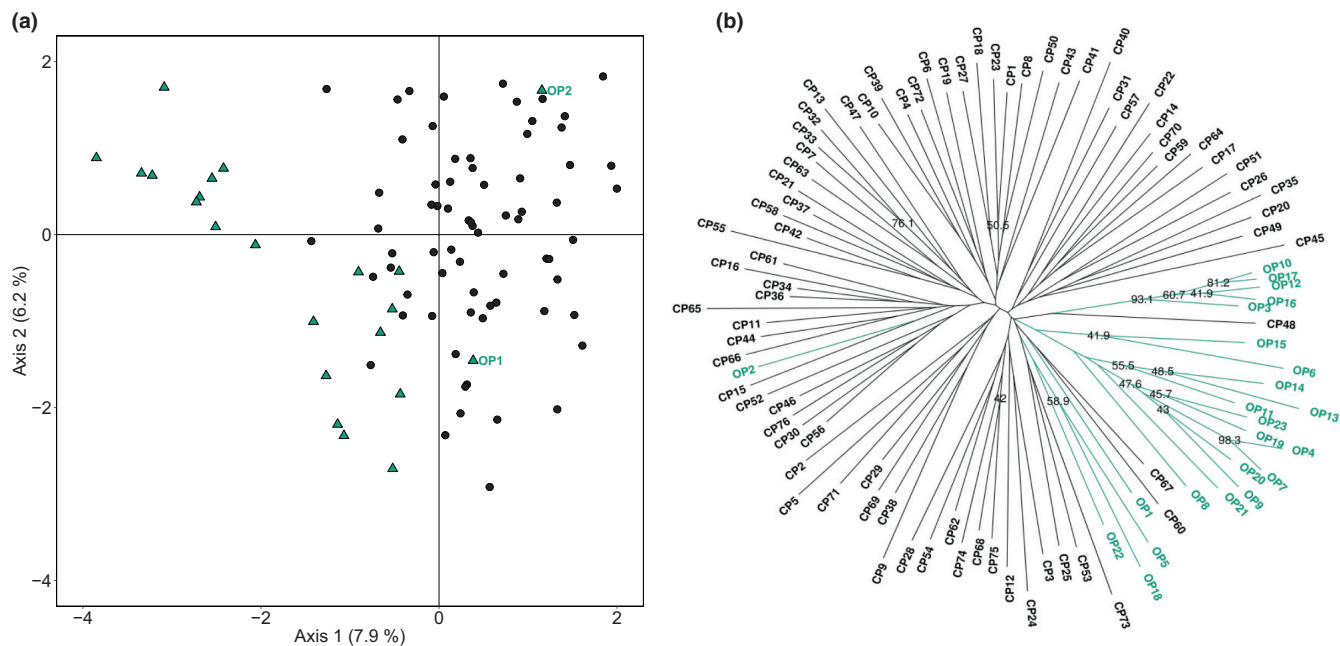


FIGURE 2 Genetic relationships between obligately parthenogenetic (OP) and cyclically parthenogenetic (CP) lineages of the clover race of *Acyrthosiphon pisum*. (a) Principal component analysis based on the 99 *A. pisum* lineages genotyped at 15 microsatellite loci. (b) Neighbour-joining tree based on shared allele distance (D_{AS}) between 99 genotypes over 15 microsatellite loci. The numbers at the nodes are bootstrap percentages (1000 replicates). Bootstrap values under 40% are not shown. OP lineages are in green and CP lineages are in black.

TABLE 1 Production of sexual and parthenogenetic morphs by obligately parthenogenetic lineages (OP) and cyclically parthenogenetic lineages (CP) of *Acyrthosiphon pisum*.

Variables	OP lineages ($n = 23$) ^a	CP lineages ($n = 76$) ^a
Total fecundity ^b	57.0	64.1
Number of sexual females	0	41.4
Number of males	2.3	7.9
Number of parthenogenetic females	54.7	14.8
% of males ^c	4.6	17.0
% of lineages that produce males	60.9	80.3

^a2–3 replicates were performed per lineage.

^bMeasured per parthenogenetic female.

^c% Measured per replicate and then averaged.

than by CP males (Tables 2 and 3, Figure 3b). The effect of the reproductive mode of the male lineage on the number of eggs laid also varied with the female lineage ($p = .048$): CP4 females were more fecund than CP10 females, but both laid less eggs when paired with OP males than with CP males (Table 3). No effect of the location where the experiments took place was observed ($p = .14$). The contribution of both fixed and random effects to the variance explained by the model was 45.4% (given by conditional R^2). This value partitioned into 14.8% due to fixed effects kept in the final model (i.e. reproductive mode, female lineage and the interaction between these two variables, given by marginal R^2) and 30.6% to random effects (i.e. male lineage, obtained by subtracting the marginal R^2 from the conditional R^2).

Among the 242 replicates in which eggs were laid, the number of fertilized eggs ranged from zero to 95 eggs per replicate. The proportion of fertilized eggs was not significantly influenced by the reproductive mode of the male lineage (GLMM, $p = .17$, Table 2, Figure 3c), with this proportion on replicates involving CP males being on average only slightly larger than that involving OP males (Table 3). We detected no significant effect of the interaction between the reproductive mode of the male lineage and the female lineage ($p = .60$), nor of the female lineage ($p = .18$) on the proportion of fertilized eggs. Experiments conducted in the laboratory showed a significantly higher proportion of fertilized eggs (48.6%) than those conducted outside the laboratory (39.5%, $p = .001$). The contribution of both fixed and random effects to the variance explained by the model was 27.9% (given by conditional R^2). This value partitioned into 3.3% due to fixed effects (i.e. location, given by marginal R^2) and 24.6% to random effects (i.e. male lineage and observation-level random effects, obtained by subtracting the marginal R^2 from the conditional R^2).

3.4 | Male reproductive success under male–male competition

After 7 days of contact between the 10 males and the five females, 483 eggs were found in the six replicates (corresponding to an average of 16.1 eggs laid per female). A total of 430 eggs (out of 483) were fertilized, this number varying from 51 to 92 depending on replicate. Except for four fertilized eggs for which paternity could not be inferred due to DNA extraction failures, all subsequent analyses

TABLE 2 Significance of factors and covariates included in the generalized linear mixed models (GLMM, see equation 1-4 in text) explaining the proportion of males produced, the number of eggs laid and the proportion of fertilized eggs without competition, and the proportion of fertilized eggs under male–male competition.

Variables	Proportion of males produced ^a			Reproductive success without competition			With competition					
	Proportion of males produced ^a			Number of eggs laid ^b			Proportion of fertilized eggs ^a					
	Estimate	<i>p</i>	<i>p</i>	Estimate	<i>p</i>	<i>p</i>	Estimate	<i>p</i>	<i>p</i>			
Factors and covariates	(see Equation 1)			(see Equation 2)			(see Equation 3)			(see Equation 4)		
Intercept	-3.864	<10 ⁻¹⁵	***	2.997	<10 ⁻¹⁵	***	-1.268	.009	**	-2.434	<10 ⁻⁶	***
Reproductive mode	-1.670	.015	*	-.918	.005	**	-1.097	.171	NS	-.608	.359	NS
Female lineage	-	-	-	.245	.150	NS	.481	.180	NS	-	-	-
Rep. mode × Fem. lin.	-	-	-	.509	.048	*	-.384	.599	NS	-	-	-
Location of the experiments	-	-	-	.200	.140	NS	1.237	.001	**	.075	.778	NS

Note: Estimates are the model coefficients; (-) means that the factor or covariate was not included in the model. The reference categories for the reproductive mode, female lineage, the interaction rep. mode × fem. lineage and location factors are respectively OP reproductive mode, CP4 lineage, OP × CP4 interaction and laboratory. Only significant terms were left in the final model. ****p* < .01; ***p* < .05; NS: *p* ≥ .05.

^aBinomial family distribution.

^bNegative binomial family distribution.

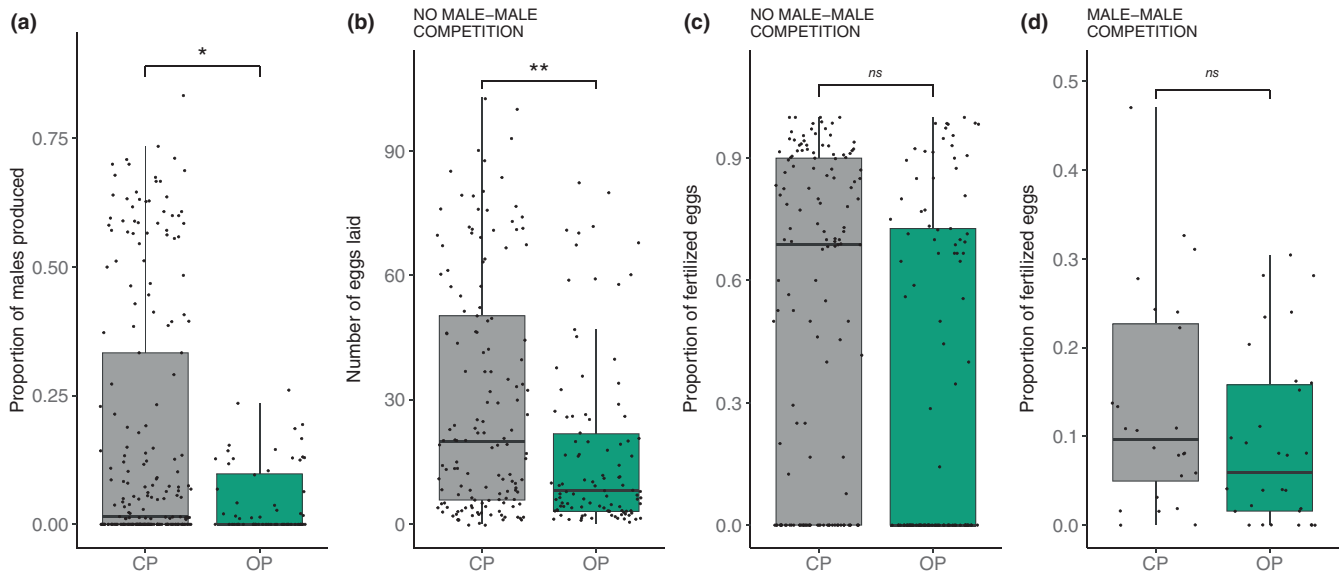


FIGURE 3 Male production and male reproductive success of obligately parthenogenetic (OP) and cyclically parthenogenetic (CP) lineages of *Acyrthosiphon pisum*. (a) Proportion of males produced per female, according to the reproductive mode of the lineage ($n = 218$ and $n = 64$ for the 76 CP and the 23 OP lineages, respectively). Reproductive success of males produced by OP and CP lineages in the absence of competition (b and c) and with competition (d). (b) Total number of eggs laid ($n = 246$), and (c) proportion of fertilized eggs for the subset of replicates in which eggs were laid ($n = 242$). (d) Proportion of eggs fertilized by one male lineage out of all fertilized eggs in mating experiments with male-male competition. Statistical significance was evaluated with generalized linear mixed models (GLMM) and is indicated by asterisks (**: $p < .01$; *: $p < .05$; ns: $p \geq .05$).

Variables	Males of OP lineages ($n = 12$) ^a	Males of CP lineages ($n = 12$) ^a
Average number of eggs laid per female ^b	3.41	6.00
Average number of eggs laid per CP4 female	4.49	6.88
Average number of eggs laid per CP10 female	2.28	5.15
% of replicates with fertilized eggs	47.0%	66.7%
Average proportion of fertilized eggs	0.36	0.51

^aFrom 1 to 27 replicates were performed per lineage.

^bConsidering the two types of female jointly (CP4 and CP10).

TABLE 3 Reproductive success of males produced by obligately parthenogenetic lineages (OP) and cyclically parthenogenetic lineages (CP) of *Acyrthosiphon pisum* in the absence of competition.

were performed on 426 eggs for which the father was identified (Table S4). From six up to nine male lineages (7.5 in average) reproduced among the six replicates (that each involved 9 different male lineages). No significant difference was observed in the proportion of eggs fertilized by males according to the reproductive mode of their lineage (GLMM, $p = .36$, Table 2): CP males sired on average 11.95% of the fertilized eggs and OP males 8.05% of them (Figure 3d). We also found no effect of the location (in the laboratory vs outside the laboratory) of the mating experiments on the proportion of fertilized eggs ($p = .78$, Table 2). The contribution of random effects (i.e. male lineage and observation-level random effects) to the variance explained by the model was 18.8%.

3.5 | Male reproductive success with and without competition

Male mating success (measured here as the number of fertilized eggs) with and without male-male competition correlated positively

but not significantly for the nine male lineages that were used in both experiments (Spearman rank correlation $r_s = 0.51$, $n = 9$, $p = .16$, Figure S2). The average number of fertilized eggs per tube was 71.7 in the experiments with male-male competition (hence each male fertilized an average of 7.17 eggs under competition) and 19.1 without competition (which corresponded to the number of eggs fertilized by single males). The increased female reproductive success in the presence of many males (the number of females was constant between the two experiments) indicated that male resource was a limiting factor for female reproduction.

4 | DISCUSSION

Here, we showed that sexual (CP) and asexual (OP) lineages of the clover host race of the pea aphid complex are geographically well separated. We then investigated the fate of sexual traits in a subset of these obligately asexual lineages, which retain male function despite low chance of encountering sexual females. OP lineages

showed a significant reduction of the production of males compared to CP lineages but the males produced were still largely functional. We detected, however, a small decline of OP male fitness that was significant only for the number of eggs laid.

4.1 | OP and CP lineages are well separated geographically

Under controlled sex-inducing conditions mimicking the shortened photoperiod in autumn, all the phenotyped lineages of *A. pisum* from eastern France—a region characterized by cold winters—produced sexual females (which lay cold resistant eggs). In contrast, none from the west (where winters are milder) produced sexual females, except for one. This demonstrates geographical segregation of OP and CP reproductive phenotypes by local adaptation to winter climate. Population genetic data also supported this conclusion: the low genotypic diversity and high proportion of clonal copies in the western pea aphid populations are consistent with predominant asexual reproduction, whereas the eastern populations consisted almost exclusively of different genotypes as predicted under sexual reproduction (Halkett et al., 2005). The few eastern repeated genotypes were found in the same field, probably reflecting clonal propagation of CP lineages in spring, not obligate parthenogenesis. The genetic tree configuration further supported the laboratory-assessed reproductive phenotype: shallow branching between several OP lineages indicated that they diversified clonally after an ancestral sex loss event, whereas CP lineages were characterized by much deeper branching.

Whether the artificial sex-inducing conditions reflect what happens in nature in autumn is absolutely crucial. We are not aware of any studies that compared the production of pea aphid sexual morphs by the same clone under artificial and natural conditions, and we have no field data demonstrating that the OP lineages do produce males in the wild. Nevertheless, the congruence between the laboratory-assessed reproductive phenotype and population genetic data strongly supports that the artificial sex-inducing conditions mirror what happens in the wild, at least for the production of sexual females. This has ecological and evolutionary consequences: first, the phenotype of western and eastern populations drastically differs in this host race of the pea aphid. Such climate-dependent segregation of aphid reproductive phenotypes has also been found in other species (Dedryver et al., 2001; Simon et al., 1999; Vorburger et al., 2003), and could affect the communities of predators and parasitoids as a result of the difference in availability of aphids as prey or hosts during winter (e.g. Le Ralec et al., 2010). Second, males produced by western OP lineages should rarely encounter sexual females. Indeed, only one of the 24 western MLGs has been determined to be a CP lineage, and when clonal copies are taken into account, CP individuals would represent about 1% of all individuals collected in the west. This geographic isolation between CP and OP lineages is potentially exacerbated in the clover race of *A. pisum* because of the limited dispersal abilities of the sexual morphs

(sexual females are all wingless and males are mostly wingless, Frantz et al., 2010), though parthenogenetic females can be winged. Therefore, this system allows the hypothesis that a large reduction of reproductive opportunities could lead to a degeneration of male sexual traits in asexual lineages to be tested.

4.2 | OP lineages produce fewer males than CP lineages

We first showed that OP lineages produced significantly fewer males than CP lineages in response to sex-inducing conditions. This reduction in male production may be the result of selection against a trait that a shift to obligate parthenogenesis renders useless and costly in the absence of sexual partners. Indeed, the production of offspring is energy-constrained, with trade-offs between the number of sexual and asexual morphs produced in aphids (Nespolo et al., 2009). In OP lineages, we did observe a reallocation of investment in males to parthenogenetic females, as male production decreased while the number of parthenogenetic females increased compared to CP lineages. Maternal control of sex determination in aphids (based on the elimination of an X chromosome to produce a male) makes this system particularly prone to rapid sex ratio evolution because there are no longer Mendelian constraints to bypass as in XX/XY or ZZ/ZW species. Nevertheless, OP lineages produced slightly fewer offspring than the CP lineages. Whether this is due to different costs in producing male or female embryos is unknown (we do not know if weight at birth differs between males and females since larvae cannot be sexed at birth). Alternatively, OP lineages could suffer from an overall reduction in fitness because the lack of recombination prevents the purging of deleterious mutations (e.g. Lynch et al., 1993).

4.3 | Males produced by OP lineages maintain good reproductive performance

In contrast, we found very little evidence for reduced functionality of males. First, a large majority of OP males fertilized eggs, demonstrating that male fertility is maintained. Second, only the number of eggs laid by a female in the presence of OP males was significantly reduced, with a small effect size. All other measures showed no difference in reproductive performance between OP and CP males, even under strong male–male competition where one would expect an exacerbation of differences, if any (e.g. Schwander et al., 2013). The reduction of the propensity of the females to lay eggs in the presence of OP males is not straightforward to interpret. As we observed that contact/mating with males seems to trigger egg laying, the number of eggs laid might be indicative of the propensity of a female to lay her eggs in the presence of a male. It can be expected that a male with altered reproductive functions will not be as stimulating to the female as a male whose same functions are under strong selection because of their importance on fitness.

4.4 | Different evolutionary trajectories for costly and neutral traits following loss of sex

The reduction in male production by OP lineages while maintaining their overall reproductive capacity is probably related to the selective mechanisms at play. Male production by OP lineages represents a net cost in the absence of female partners and therefore should be rapidly counter-selected, whereas male degeneration occurs only through relaxed selection. This supports predictions and observations that counter-selected sexual traits degenerate more rapidly than traits evolving under relaxed selection (van der Kooi & Schwander, 2014).

Yet, male production by OP lineages is a common pattern in aphids (Blackman, 1974; Dedryver et al., 1998; Frantz et al., 2006; Simon et al., 1991) and also occurs at least occasionally in other asexual species such as stick insects (Schwander et al., 2013), snails (Neiman et al., 2012), thrips (Mirab-Balou & Chen, 2010), *Daphnia* (Innes & Hebert, 1988), darwinulid ostracods (Smith et al., 2006) or brine shrimps (MacDonald & Browne, 1987). Different hypotheses can be invoked to explain why more than half of our OP lineages still produced males and why these were largely functional.

First, sex loss might be relatively recent, in which case OP lineages would have spent only a short time period under relaxed selection on male-related traits. Indeed, the neutral decay of male traits in asexual lineages could require thousands of years to result in significant phenotypic differences, a timescale rarely reached by asexual lineages due to mutational meltdown (Schurko et al., 2009). Interestingly, the genetic tree based on microsatellite markers revealed that many OP lineages clustered together on a well-supported branch. This suggests that a sex loss event in this branch happened sufficiently long ago for subsequent clonal diversification. Whether it is old enough to expect a degradation of male traits remains an open question, since we could not date this event. Two other OP lineages each branched with CP lineages and grouped with the cluster of CP lineages in the PCA. They could represent independent and possibly more recent sex loss events. Interestingly, these two lineages were also those showing - among OP lineages - the highest ability to induce female egg laying. Hence, OP lineages likely lost the ability to produce sexual females at different times as indicated by their possible different origins, which could influence the extent of male-related trait decay between lineages.

Second, gene flow between OP and CP lineages is another factor that could prevent male degeneration in OP lineages. Although eastern and western populations are characterized by contrasted phenotypes, the two types of lineages are likely to coexist in intermediate climate habitats (Dedryver et al., 2001; Rispe et al., 1998; Rispe & Pierre, 1998) or simply through dispersal to suboptimal habitats. Mating between OP males and CP females would introduce *op* allele(s) in a new genetic background. Depending on the dominance of *op* allele(s), new OP lineages could be generated in this or subsequent generations. Thus, these new (and young) OP lineages could possess allelic versions from CP lineages for genes for male production and male fitness, which should therefore be fully functional.

Population genomic data on OP and CP lineages of the alfalfa host race of the pea aphid demonstrated the recessivity of the *op* allele at the locus associated with sex loss and low genome wide differentiation outside this locus ($F_{ST}=2.4\%$, Jaquiéry et al., 2014). This suggests sufficient gene flow between OP and CP populations to counterbalance drift. The slightly higher F_{ST} (4.8%) in the clover host race, which was studied here, does not rule out the possibility of gene exchange. Finally, the male functionality in OP lineages may be maintained by pleiotropy, where the same gene network is required in other processes for parthenogenetic females. Pleiotropic constraints may increase the complexity of selection on genes important to males (Mank et al., 2008).

4.5 | Origin and number of sex loss events

The low branch resolution prevents assertion of whether the three main clusters of OP lineages in the D_{AS} tree represent independent or shared sex loss events. Thus, we cannot rule out that the majority of our lineages derive from the same transition to obligate asexuality, which would imply that each OP lineage should not be considered a statistically independent observation. Yet, another lineage (OP2) clearly clustered with CP lineages suggesting that transitions to obligate parthenogenesis could have occurred at least twice in this host race. This finding relies heavily on the OP2 lineage, so it is crucial to be certain of its phenotype. This lineage was determined as OP under sex-inducing conditions in two independent experiments and a total of 6 individual copies of this clone were found in clover fields 2 km apart, a pattern that also supports asexual reproduction. Further sampling of wild individuals is necessary to get a more comprehensive picture of the frequency of such sex loss events. Also, our data does not inform on how sex is lost. New OP aphid lineages have been shown to originate from multiple genetic mechanisms depending on species: (i) de novo mutation(s) at the gene(s) controlling the production of sexual morphs (Delmotte et al., 2001), (ii) contagious asexuality resulting from crosses between OP males and CP sexual females (Halkett et al., 2008; Jaquiéry et al., 2014), (iii) hybridization between closely related species (Delmotte et al., 2003). The ancestor of OP2 could have lost the ability to produce sexual females by spontaneous mutation or by contagious asexuality, the resolution of our tree being insufficient to disentangle the two hypotheses, especially if the *op* allele(s) is (are) recessive as in the alfalfa host race of the pea aphid complex (Jaquiéry et al., 2014). The third mechanism is unlikely in our system as the OP and CP lineages do not differ in heterozygosity, as would have been the case if OP lineages had been generated by interspecific crosses.

5 | CONCLUSION

This study shows that the transition from cyclical to obligate parthenogenesis results in changes in selective regimes between CP

and OP populations with various evolutionary consequences: they manifest as (1) a reduction in the proportion of males produced by OP lineages, supposedly resulting from counter-selection, and (2) a maintenance of egg fertilizing capacity of OP males, supposedly resulting from slow drift-driven decay or underestimated reproductive opportunities. Sampling in more intermediate habitats would help to determine the extent of distribution overlap between the OP and CP lineages and better characterize potential gene flow. Other major issues in this study are whether and how quickly a trait will decay when selection exerted on it is removed or relaxed, and how this will relate to the trait functions and the mechanisms involved. Our study contributes to show that species adaptation following a change in reproductive system also involves the decay of useless or maladaptive traits. To expand this analysis to other functional traits, future studies should consider the evolutionary consequences of sex loss at the transcriptomic and genomic scale. This would enable us to better apprehend and quantify the decay or maintenance of sexual traits under asexuality, and especially those related to sexual females, that are not exposed to selection in OP aphids.

AUTHOR CONTRIBUTIONS

JCS, JJ, YO and HD conceived and designed the experiments. HD, JJ, MR, FM, RC and GD performed the experiments. HD, JJ, JCS, LM and YO analysed the data. HD, JJ and JCS wrote the paper.

ACKNOWLEDGEMENTS

We warmly thank Editor Tatiana Giraud and three anonymous referees for constructive comments on previous drafts, as well as Chris Bass for his critical reading.

FUNDING INFORMATION

This work was supported by grants from the French Research Agency (ANR) SexAphid (ANR-09-GENM-017-001) and Mecadapt (ANR-11-BSV7-005-01), the INRAE-SPE Department (AAP GenAsex and half a PhD grant for HD) and the Region Bretagne (half a PhD grant for HD).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicting interest.

DATA AVAILABILITY STATEMENT

The authors confirm that all relevant data underlying the findings are fully available without restriction, within the paper and its supporting information files. Relevant data are fully available in Tables S1 to S4.

ORCID

Hélène Defendini  <https://orcid.org/0000-0002-1016-3677>

Julie Jaquiéry  <https://orcid.org/0000-0002-7394-0883>

REFERENCES

Adachi-Hagimori, T., & Miura, K. (2020). Limited mating ability of a wasp strain with *rickettsia*-induced thelytoky. *Annals of the Entomological*

- Society of America*, 113(5), 355–358. <https://doi.org/10.1093/aesa/saaa007>
- Bailleul, D., Stoeckel, S., & Arnaud-Haond, S. (2016). RClone: a package to identify MultiLocus clonal lineages and handle clonal data sets in R. *Methods in Ecology and Evolution*, 7(8), 966–970. <https://doi.org/10.1111/2041-210X.12550>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bell, G. (1982). *The masterpiece of nature: The evolution and genetics of sexuality*. University of California Press.
- Blackman, R. L. (1972). The inheritance of life-cycle differences in *Myzus persicae* (Sulz.) (Hom., Aphididae). *Bulletin of Entomological Research*. Cambridge University Press, 62(2), 281–294. <https://doi.org/10.1017/S0007485300047726>
- Blackman, R. L. (1974). Life-cycle variation of *Myzus persicae* (Sulz.) (Hom., Aphididae) in different parts of the world, in relation to genotype and environment. *Bulletin of Entomological Research*, 63(4), 595–607. <https://doi.org/10.1017/S0007485300047830>
- Blackman, R. L. (1975). Photoperiodic determination of the male and female sexual morphs of *Myzus persicae*. *Journal of Insect Physiology*, 21(2), 435–453. [https://doi.org/10.1016/0022-1910\(75\)90036-0](https://doi.org/10.1016/0022-1910(75)90036-0)
- Blackman, R. L. (1987). Reproduction, cytogenetics and development. In A. K. Minks & P. Harrewijn (Eds.), *Aphids: Their biology, natural enemies and control* (Vol. 2A, pp. 163–195). Elsevier.
- Bougeard, S., & Dray, S. (2018). Supervised multiblock analysis in R with the ade4 package. *Journal of Statistical Software*, 86, 1–17. <https://doi.org/10.18637/jss.v086.i01>
- Caillaud, M. C., Mondor-Genson, G., Levine-Wilkinson, S., Mieuze, L., Frantz, A., Simon, J. C., & Coeur D'acier, A. (2004). Microsatellite DNA markers for the pea aphid *Acyrtosiphon pisum*. *Molecular Ecology Notes*, 4(3), 446–448. <https://doi.org/10.1111/j.1471-8286.2004.00676.x>
- Chakraborty, R., & Jin, L. (1993). A unified approach to study hyper-variable polymorphisms: Statistical considerations of determining relatedness and population distances. In S. D. J. Pena, R. Chakraborty, J. T. Epplen, & A. J. Jeffreys (Eds.), *DNA fingerprinting: State of the science* (pp. 153–175). Birkhäuser Verlag. https://doi.org/10.1007/978-3-0348-8583-6_14
- Chasnov, J. R., & Chow, K. L. (2002). Why are there males in the hermaphroditic species *Caenorhabditis elegans*? *Genetics*, 160(3), 983–994. <https://doi.org/10.1093/genetics/160.3.983>
- Dedryver, C.-A., Hullé, M., Le Gallic, J.-F., Caillaud, M. C., & Simon, J.-C. (2001). Coexistence in space and time of sexual and asexual populations of the cereal aphid *Sitobion avenae*. *Oecologia*, 128(3), 379–388. <https://doi.org/10.1007/s004420100674>
- Dedryver, C. A., Le Gallic, J. F., Gauthier, J. P., & Simon, J. C. (1998). Life cycle of the cereal aphid *Sitobion avenae* F: Polymorphism and comparison of life history traits associated with sexuality. *Ecological Entomology*, 23(2), 123–132.
- Dedryver, C.-A., Le Gallic, J.-F., Mahéo, F., Parisey, N., & Tagu, D. (2012). Delayed setting of the photoperiodic response in recombinant clones of the aphid species *Sitobion avenae*. *Ecological Entomology*, 37(4), 293–299. <https://doi.org/10.1111/j.1365-2311.2012.01365.x>
- Dedryver, C.-A., Le Gallic, J.-F., Mahéo, F., Simon, J.-C., & Dedryver, F. (2013). The genetics of obligate parthenogenesis in an aphid species and its consequences for the maintenance of alternative reproductive modes. *Heredity*, 110(1), 39–45. <https://doi.org/10.1038/hdy.2012.57>
- Delmotte, F., Leterme, N., Bonhomme, J., Rispe, C., & Simon, J.-C. (2001). Multiple routes to asexuality in an aphid species. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 268(1483), 2291–2299. <https://doi.org/10.1098/rspb.2001.1778>

- Delmotte, F., Sabater-Muñoz, B., Prunier-Leterme, N., Latorre, A., Sunnucks, P., Rispe, C., & Simon, J.-C. (2003). Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. *Evolution*, 57(6), 1291–1303. <https://doi.org/10.1111/j.0014-3820.2003.tb00337.x>
- Dixon, A. F. G. (1998). *Aphid ecology* (2nd ed.). Chapman & Hall.
- Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. SAGE Publications.
- Frantz, A., Plantegenest, M., & Simon, J.-C. (2006). Temporal habitat variability and the maintenance of sex in host populations of the pea aphid. *Proceedings of the Royal Society B: Biological Sciences*, 273(1603), 2887–2891. <https://doi.org/10.1098/rspb.2006.3665>
- Frantz, A., Plantegenest, M., & Simon, J.-C. (2010). Host races of the pea aphid *Acyrtosiphon pisum* differ in male wing phenotypes. *Bulletin of Entomological Research*, 100(1), 59–66. <https://doi.org/10.1017/S0007485309006750>
- Goodenough, U., & Heitman, J. (2014). Origins of eukaryotic sexual reproduction. *Cold Spring Harbor Perspectives in Biology*, 6(3), a016154. <https://doi.org/10.1101/cshperspect.a016154>
- Gottlieb, Y., & Zchori-Fein, E. (2001). Irreversible thelytokous reproduction in *Muscidifurax uniraptor*. *Entomologia Experimentalis et Applicata*, 100(3), 271–278. <https://doi.org/10.1046/j.1570-7458.2001.00874.x>
- Goudet, J., & Jombart, T. (2022). *Hierfstat: Estimation and tests of hierarchical F-statistics*. R Package Version 0.5-11. CRAN.R-project. <https://CRAN.R-project.org/package=hierfstat>
- Grimaldi, D., & Engel, M. S. (2005). *Evolution of the insects*. Cambridge University Press.
- Hales, D. F., Tomiuk, J., Woehrmann, K., & Sunnucks, P. (1997). Evolutionary and genetic aspects of aphid biology: A review. *EJE*, 94(1), 1–55.
- Halkett, F., Plantegenest, M., Bonhomme, J., & Simon, J.-C. (2008). Gene flow between sexual and facultatively asexual lineages of an aphid species and the maintenance of reproductive mode variation. *Molecular Ecology*, 17(12), 2998–3007. <https://doi.org/10.1111/j.1365-294X.2008.03798.x>
- Halkett, F., Simon, J.-C., & Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, 20(4), 194–201. <https://doi.org/10.1016/j.tree.2005.01.001>
- Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, 2, e616. <https://doi.org/10.7717/peerj.616>
- Harrison, X. A. (2015). A comparison of observation-level random effect and Beta-binomial models for modelling overdispersion in binomial data in ecology & evolution. *PeerJ*, 3, e1114. <https://doi.org/10.7717/peerj.1114>
- Innes, D. J., Fox, C. J., & Winsor, G. L. (2000). Avoiding the cost of males in obligately asexual *Daphnia pulex* (Leydig). *Proceedings of the Royal Society of London Series B: Biological Sciences*, 267(1447), 991–997. <https://doi.org/10.1098/rspb.2000.1101>
- Innes, D. J., & Hebert, P. D. N. (1988). The origin and genetic basis of obligate parthenogenesis in *Daphnia pulex*. *Evolution*, 42(5), 1024–1035. <https://doi.org/10.1111/j.1558-5646.1988.tb02521.x>
- Jalinsky, J., Logsdon, J. M., & Neiman, M. (2020). Male phenotypes in a female framework: Evidence for degeneration in sperm produced by male snails from asexual lineages. *Journal of Evolutionary Biology*, 33(8), 1050–1059. <https://doi.org/10.1111/jeb.13632>
- Jaquière, J., Stoeckel, S., Larose, C., Nouhaud, P., Rispe, C., Mieuze, L., Bonhomme, J., Mahéo, F., Legeai, F., Gauthier, J. P., Prunier-Leterme, N., Tagu, D., & Simon, J.-C. (2014). Genetic control of contagious asexuality in the pea aphid. *PLoS Genetics*, 10(12), e1004838. <https://doi.org/10.1371/journal.pgen.1004838>
- Jaquière, J., Stoeckel, S., Rispe, C., Mieuze, L., Legeai, F., & Simon, J.-C. (2012). Accelerated evolution of sex chromosomes in aphids, an XO system. *Molecular Biology and Evolution*, 29(2), 837–847. <https://doi.org/10.1093/molbev/msr252>
- Johnston, M. O., Das, B., & Hoeh, W. R. (1998). Negative correlation between male allocation and rate of self-fertilization in a hermaphroditic animal. *Proceedings of the National Academy of Sciences*, 95(2), 617–620. <https://doi.org/10.1073/pnas.95.2.617>
- Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Kampfraath, A. A., Dudink, T. P., Kraaijeveld, K., Ellers, J., & Zizzari, Z. V. (2020). Male sexual trait decay in two asexual springtail populations follows neutral mutation accumulation theory. *Evolutionary Biology*, 47(4), 285–292. <https://doi.org/10.1007/s11692-020-09511-z>
- Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6, 208. <https://doi.org/10.3389/fgene.2015.00208>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>
- Kraaijeveld, K., Anvar, S. Y., Frank, J., Schmitz, A., Bast, J., Wilbrandt, J., Petersen, M., Ziesmann, T., Niehuis, O., de Knijff, P., den Dunnen, J. T., & Ellers, J. (2016). Decay of sexual trait genes in an asexual parasitoid wasp. *Genome Biology and Evolution*, 8(12), 3685–3695. <https://doi.org/10.1093/gbe/evw273>
- Le Ralec, A., Anselme, C., Outreman, Y., Poirié, M., van Baaren, J., Le Lann, C., & van Alphen, J. J.-M. (2010). Evolutionary ecology of the interactions between aphids and their parasitoids. *Comptes Rendus Biologies*, 333(6–7), 554–565. <https://doi.org/10.1016/j.crv.2010.03.010>
- Le Trionnaire, G., Francis, F., Jaubert-Possamai, S., Bonhomme, J., De Pauw, E., Gauthier, J.-P., & Tagu, D. (2009). Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid. *BMC Genomics*, 10(1), 456. <https://doi.org/10.1186/1471-2164-10-456>
- Loxdale, H. D., Hardie, J., Halbert, S., Footitt, R., Kidd, N. A. C., & Carter, C. I. (1993). The relative importance of short- and long-range movement of flying aphids. *Biological Reviews*, 68(2), 291–311. <https://doi.org/10.1111/j.1469-185X.1993.tb00998.x>
- Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139. <https://doi.org/10.21105/joss.03139>
- Lynch, M., Bürger, R., Butcher, D., & Gabriel, W. (1993). The mutational meltdown in asexual populations. *Journal of Heredity*, 84(5), 339–344. <https://doi.org/10.1093/oxfordjournals.jhered.a111354>
- MacDonald, G. H., & Browne, R. A. (1987). Inheritance and reproductive role of rare males in a parthenogenetic population of the brine shrimp. *Artemia Parthenogenetica*. *Genetica*, 75(1), 47–53. <https://doi.org/10.1007/BF00056032>
- Mank, J. E., Hultin-Rosenberg, L., Zwahlen, M., & Ellegren, H. (2008). Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *The American Naturalist*, 171(1), 35–43. <https://doi.org/10.1086/523954>
- Mirab-Balou, M., & Chen, X. X. (2010). First description of the male of the wheat thrips, *Anaphothrips obscurus* (Thysanoptera: Thripidae). *Zootaxa*, 2540(1), 65–68. <https://doi.org/10.11646/zootaxa.2540.1.4>
- Miura, T., Braendle, C., Shingleton, A., Sisk, G., Kambhampati, S., & Stern, D. L. (2003). A comparison of parthenogenetic and sexual embryogenesis of the pea aphid *Acyrtosiphon pisum* (Hemiptera: Aphidoidea). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 295B(1), 59–81. <https://doi.org/10.1002/jez.b.3>

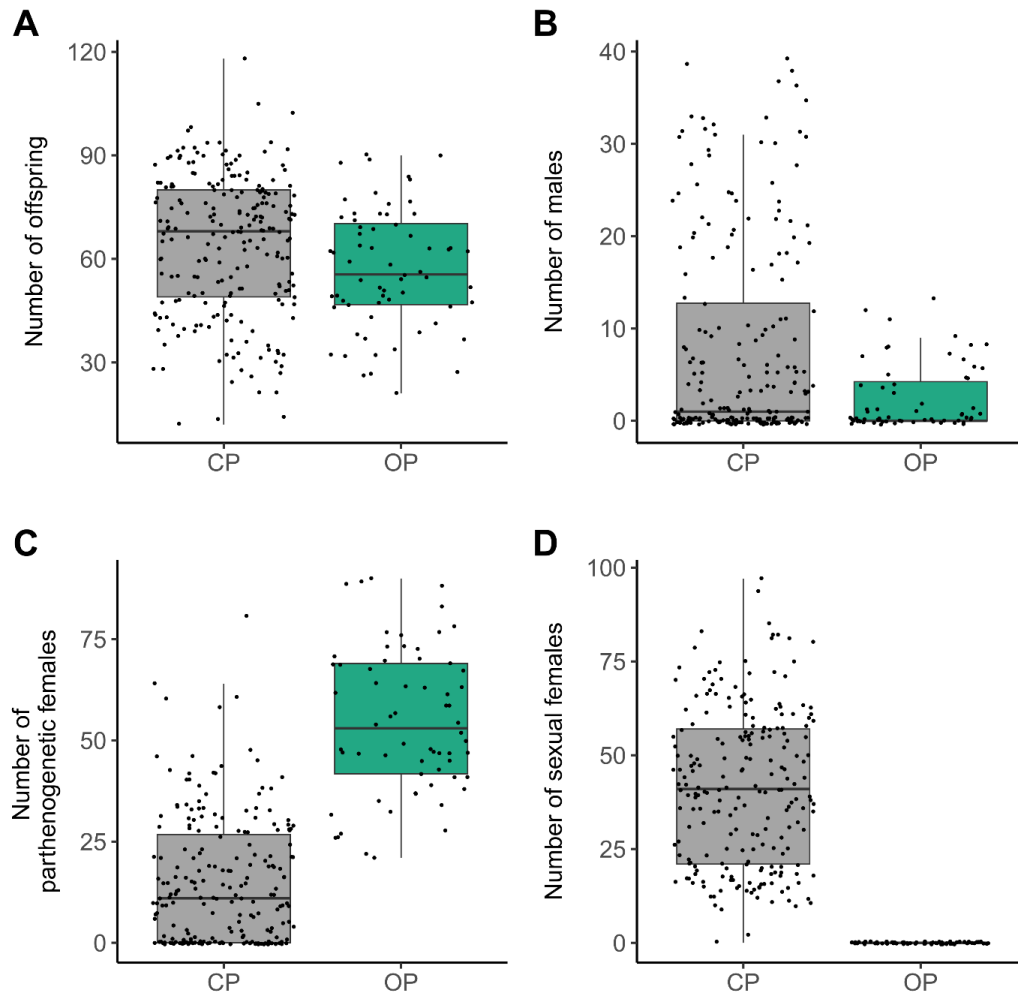
- Miyazaki, M. (1987). Forms and morphs of aphids. In A. K. Minks & P. Harrewijn (Eds.), *Aphids, their biology, natural enemies and control* (pp. 163–195). Elsevier.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Neiman, M., Larkin, K., Thompson, A. R., & Wilton, P. (2012). Male offspring production by asexual *Potamopyrgus antipodarum*, a New Zealand snail. *Heredity*, 109(1), 57–62. <https://doi.org/10.1038/hdy.2012.13>
- Nespolo, R. F., Halkett, F., Figueroa, C. C., Plantegenest, M., & Simon, J.-C. (2009). Evolution of trade-offs between sexual and asexual phases and the role of reproductive plasticity in the genetic architecture of aphid life histories. *Evolution*, 63(9), 2402–2412. <https://doi.org/10.1111/j.1558-5646.2009.00706.x>
- Nouhaud, P., Peccoud, J., Mahéo, F., Mieuze, L., Jaquiéry, J., & Simon, J.-C. (2014). Genomic regions repeatedly involved in divergence among plant-specialized pea aphid biotypes. *Journal of Evolutionary Biology*, 27(9), 2013–2020. <https://doi.org/10.1111/jeb.12441>
- Pannebakker, B. A., Schidlo, N. S., Boskamp, G. J. F., Dekker, L., Dooren, T. J. M. V., Beukeboom, L. W., Zwaan, B. J., Brakefield, P. M., & Alphen, J. J. M. V. (2005). Sexual functionality of *Leptopilina clavipes* (Hymenoptera: Figitidae) after reversing *Wolbachia*-induced parthenogenesis. *Journal of Evolutionary Biology*, 18(4), 1019–1028. <https://doi.org/10.1111/j.1420-9101.2005.00898.x>
- Papura, D., Simon, J.-C., Halkett, F., Delmotte, F., Le Gallic, J.-F., & Dedryver, C.-A. (2003). Predominance of sexual reproduction in Romanian populations of the aphid *Sitobion avenae* inferred from phenotypic and genetic structure. *Heredity*, 90(5), 397–404. <https://doi.org/10.1038/sj.hdy.6800262>
- Peccoud, J., Figueroa, C. C., Silva, A. X., Ramirez, C. C., 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(21), 4608–4618. <https://doi.org/10.1111/j.1365-294X.2008.03949.x>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rimbault, M., Legeai, F., Peccoud, J., Mieuze, L., Call, E., Nouhaud, P., Defendini, H., Mahéo, F., Marande, W., Théron, N., Tagu, D., Le Trionnaire, G., Simon, J.-C., & Jaquiéry, J. (2022). A quarter-million-year-old polymorphism drives reproductive mode variation in the pea aphid. *BioRxiv*. <https://doi.org/10.1101/2022.10.24.513443>
- Rispe, C., & Pierre, J.-S. (1998). Coexistence between cyclical parthenogens, obligate parthenogens, and intermediates in a fluctuating environment. *Journal of Theoretical Biology*, 195(1), 97–110. <https://doi.org/10.1006/jtbi.1998.0784>
- Rispe, C., Pierre, J.-S., Simon, J.-C., & Gouyon, P.-H. (1998). Models of sexual and asexual coexistence in aphids based on constraints. *Journal of Evolutionary Biology*, 11(6), 685–701. <https://doi.org/10.1046/j.1420-9101.1998.11060685.x>
- Sack, C., & Stern, D. L. (2007). Sex and death in the male pea aphid, *Acyrtosiphon pisum*: The life-history effects of a wing dimorphism. *Journal of Insect Science*, 7(1), 45–49. <https://doi.org/10.1673/031.007.4501>
- Schurko, A. M., Neiman, M., & Logsdon, J. M. (2009). Signs of sex: What we know and how we know it. *Trends in Ecology & Evolution*, 24(4), 208–217. <https://doi.org/10.1016/j.tree.2008.11.010>
- Schwander, T., Crespi, B. J., Gries, R., & Gries, G. (2013). Neutral and selection-driven decay of sexual traits in asexual stick insects. *Proceedings of the Royal Society B: Biological Sciences*, 280(1764), 20130823. <https://doi.org/10.1098/rspb.2013.0823>
- Simon, J.-C., Baumann, S., Sunnucks, P., Hebert, P. D. N., Pierre, J.-S., Le Gallic, J.-F., & Dedryver, C.-A. (1999). Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, 8(4), 531–545. <https://doi.org/10.1046/j.1365-294x.1999.00583.x>
- Simon, J. C., Blackman, R. L., & Gallic, J. F. (1991). Local variability in the life cycle of the bird cherry-oat aphid, *Rhopalosiphum padi* (Homoptera: Aphididae) in western France. *Bulletin of Entomological Research*, 81(3), 315–322. <https://doi.org/10.1017/S000748530003599>
- Simon, J.-C., Delmotte, F., Rispe, C., & Crease, T. (2003). Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society*, 79(1), 151–163. <https://doi.org/10.1046/j.1095-8312.2003.00175.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), 488–496. <https://doi.org/10.1016/j.crvi.2010.03.003>
- Smith, R. J., Kamiya, T., & Horne, D. J. (2006). Living males of the 'ancient asexual' *Darwinulidae* (Ostracoda: Crustacea). *Proceedings of the Royal Society B: Biological Sciences*, 273(1593), 1569–1578. <https://doi.org/10.1098/rspb.2005.3452>
- van der Kooij, C. J., & Schwander, T. (2014). On the fate of sexual traits under asexuality. *Biological Reviews*, 89(4), 805–819. <https://doi.org/10.1111/brv.12078>
- 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(12), 3493–3504. <https://doi.org/10.1046/j.1365-294X.2003.01998.x>
- Vrijenhoek, R. C., Dawley, R. M., Cole, C. J., & Bogart, J. P. (1989). A list of known unisexual vertebrates. In R. M. Dawley & J. P. Bogart (Eds.), *Evolution and ecology of unisexual vertebrates* (pp. 19–23). University of the State of New York.
- Wieczorek, K., Kanturski, M., Sempruch, C., & Świątek, P. (2019). The reproductive system of the male and oviparous female of a model organism – The pea aphid, *Acyrtosiphon pisum* (Hemiptera, Aphididae). *PeerJ*, 7, e7573. <https://doi.org/10.7717/peerj.7573>
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). Ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. <https://doi.org/10.1111/2041-210X.12628>
- Zchori-Fein, E., Roush, R. T., & Hunter, M. S. (1992). Male production induced by antibiotic treatment in *Encarsia Formosa* (Hymenoptera: Aphelinidae), an asexual species. *Experientia*, 48(1), 102–105.

SUPPORTING INFORMATION

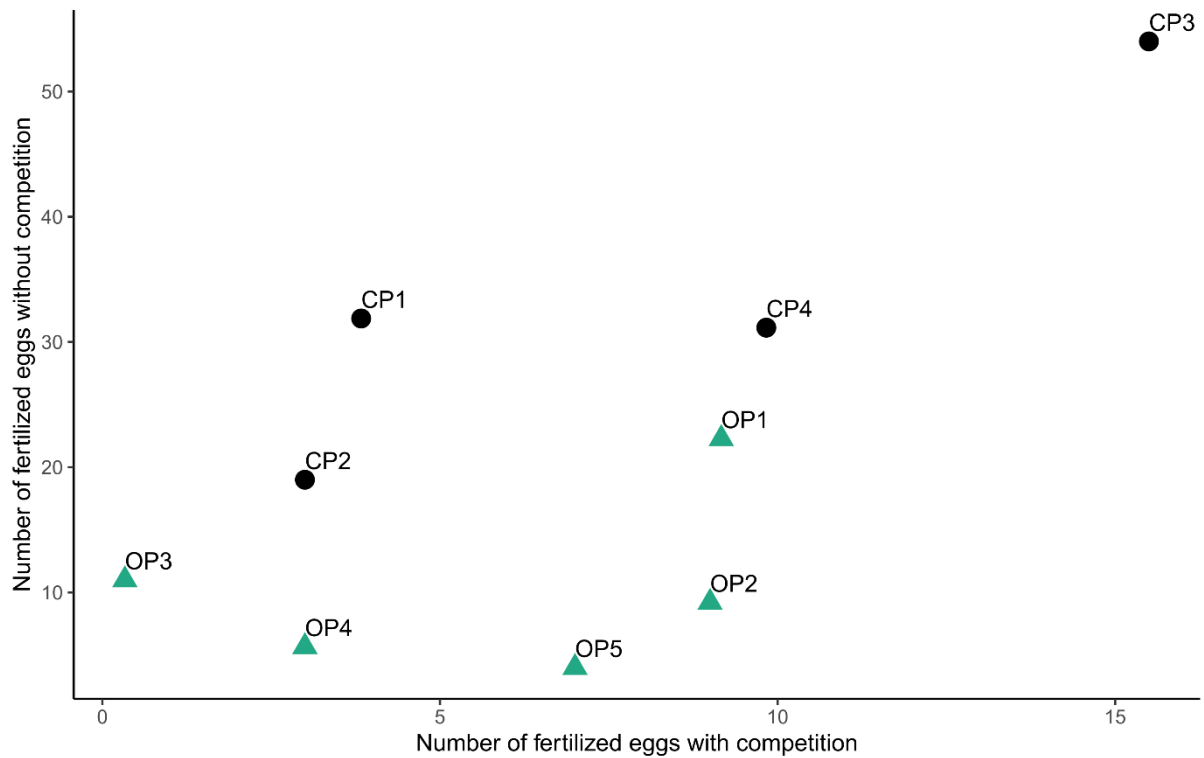
Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Defendini, H., Rimbault, M., Mahéo, F., Cloteau, R., Denis, G., Mieuze, L., Outreman, Y., Simon, J.-C., & Jaquiéry, J. (2023). Evolutionary consequences of loss of sexual reproduction on male-related traits in parthenogenetic lineages of the pea aphid. *Molecular Ecology*, 32, 3672–3685. <https://doi.org/10.1111/mec.16961>

Supplementary materials for chapter 3



Supplementary Figure S1. Offspring production of obligately parthenogenetic (OP) and cyclically parthenogenetic (CP) lineages of *Acyrthosiphon pisum*. (A) Number of total offspring (all morphs included), (B) number of males, (C) number of parthenogenetic females and (D) number of sexual females produced per female, according to the reproductive mode of the lineage ($n = 218$ and $n = 64$ for the 76 CP and the 23 OP lineages, respectively).



Supplementary Figure S2. Absolute male fertilization success in experiments with competition (x-axis, 10 males for 5 females) and without competition (y-axis, 1 male for 5 females). Cyclically parthenogenetic (CP) males are shown with black dots and obligate parthenogenetic (OP) males with green triangles. The average number of fertilized eggs per male lineage with and without competition correlated positively although not significantly for the nine lineages used in the two experiments (Spearman rank correlation $r_s = 0.51$, $n = 9$, $p = 0.16$).

LEGEND OF SUPPLEMENTARY TABLES

Supplementary Table S1. Geographic origin and genotype at 15 microsatellite markers for the 219 pea aphid individuals sampled in clover fields that established a colony in the laboratory. Clonal copies of the same multilocus genotype (MLG) are identified ("CloneID" column), the reproductive phenotype is given for the tested lineages, and the last columns indicate the lineages used for the different types of analysis as well as their simplified name used in the manuscript.

Supplementary Table S2. Number of males, sexual females and parthenogenetic females produced by 99 different lineages when placed under sex-inducing conditions.

Supplementary Table S3. Reproductive success of males from 24 different lineages in mating experiments without male-male competition.

Supplementary Table S4. Reproductive success of males in mating experiments with male-male competition.

Supplementary tables are available on
<https://onlinelibrary.wiley.com/doi/full/10.1111/mec.16961>

Chapter 4

In chapter 4, I studied whether the loss of sex in the pea aphid results in shifts in expression of morph-biased genes supposedly associated with relaxation of intra-locus sexual conflict. This manuscript will be submitted soon to Proceedings of the Royal Society. I also presented a poster about this work at the Jacques Monod conference in Roscoff (“Sex unfolded: sex, asex, sexes”) in September 2023, which is appended to the PhD manuscript (Annex 1).

The release of sexual conflict after sex loss is associated with evolutionary changes in gene expression

Hélène Defendini¹, Nathalie Prunier-Leterme¹, Stéphanie Robin², Sonia Lameiras³, Sylvain Baulande³, Jean-Christophe Simon¹, Julie Jaquiéry¹

¹ UMR 1349 IGEPP, INRAE, Institut Agro, Université de Rennes, 35653 Le Rheu, France

² UMR 1349 IGEPP, INRAE, Institut Agro, Université de Rennes, 35000 Rennes, France

³ Institut Curie, PSL University, ICGex Next-Generation Sequencing Platform, 75005 Paris, France

Corresponding author: Julie Jaquiery, Julie.Jaquiery@inrae.fr, <https://orcid.org/0000-0002-7394-0883>

Abstract

Sexual conflict is widespread because despite sharing the same genome, males and females exhibit different phenotypes. Sexually dimorphic gene expression may help reduce conflict, but the expression of many genes remains sub-optimal due to unresolved tensions between the sexes. Asexual lineages lack such conflict, making them relevant models for understanding the extent to which sexual conflict influences gene expression. We investigate the evolution of sexual conflict subsequent to sex loss by contrasting the gene expression patterns of sexual and asexual lineages in the pea aphid. Although asexual lineages of this aphid produce a small number of males in autumn, their mating opportunities are limited because of geographic isolation between sexual and asexual lineages. Therefore, gene expression in parthenogenetic females of asexual lineages is no longer constrained by that of other morphs. We found that the expression of genes in males from asexual lineages tended towards the parthenogenetic female optimum, in agreement with theoretical predictions. Surprisingly, males and parthenogenetic females of asexual lineages over-expressed genes normally found in the ovaries and testes of sexual morphs. These changes in gene expression in asexual lineages may arise from the relaxation of selection or the dysregulation of gene networks otherwise used in sexual lineages.

Keywords: sexual antagonism, parthenogenesis, asexuality, transcriptomics, sex-biased gene expression, sexual dimorphism

INTRODUCTION

Males and females share most of their genome, but the sexes may have different fitness optima for many phenotypic traits. When the genetic architecture of these traits is shared between the sexes, this can lead to intra-locus sexual conflict, where a trait is prevented from evolving towards its fitness optimum in males, females or both (Bonduriansky & Chenoweth, 2009; Lande, 1980). The evolution of sexually dimorphic gene expression may help to reduce conflict, although the expression of many genes may still be suboptimal due to remaining genetic and functional constraints (Griffin et al., 2013; Mank et al., 2008; Tosto et al., 2023). As a result, genes with a current sexually dimorphic expression should reflect different stages of resolved or partially unresolved sexual conflicts (Cheng & Kirkpatrick, 2016). The extent to which these intra-locus conflicts are resolved in natural populations remains difficult to assess because of the small effects of the presumably numerous sexually antagonistic loci distributed throughout the genome (Cheng & Kirkpatrick, 2016; Chippindale et al., 2001; Innocenti & Morrow, 2010).

One way of addressing this question is to study the evolution of the transcriptome in sexual species when selective pressures are artificially altered. Thereafter, genes with sexually dimorphic expression are referred to as sex-biased genes and further subdivided into male-biased and female-biased genes, depending on which sex shows higher expression. First, enforcing monogamy in polygamous species is expected to reduce sexual selection on males and thus relax sexually antagonistic selection (Holland & Rice, 1999). Evolution under monogamy would thus lead to a change in sex-biased gene expression in both sexes towards the female optimum, resulting in increased expression of female-biased genes and decreased expression of male-biased genes. Two studies that experimentally evolved different polygamous *Drosophila* species under monogamy for more than 100 generations had mixed results. The male and female transcriptomes of *D. melanogaster* indeed showed pattern of feminization (Hollis et al., 2014) but in *D. pseudoobscura*, shifts in the expression of sex-biased genes were mainly tissue- and condition-dependent and, contrary to predictions, transcriptomes tended to be masculinized (Veltos et al., 2017). Major limitations of these studies are that evolutionary time elapsed since the change in mating system is short and that enforcing monogamy does not completely remove selection on males, as their genes are still passed on to the next generation.

Species that have lost sex are of particular interest because intra-locus conflict between sexes no longer applies. The comparison of the transcriptomes of asexual and closely-related sexual species provides a good opportunity to study gene expression evolution after the suppression of any form of sexual antagonism following sex loss. Any asexual lineage that invests more in female functions compared to sexual species - at the expense of males that are absent and thus do not constrain gene expression in females - should be favoured by selection. Thus, we expect higher expression of female-biased genes and reduced expression of male-biased genes in asexual as compared to sexual species (Huylmans et al., 2021; Parker et al., 2019a).

Parker et al. (2019a) compared expression of sex-biased genes in five pairs of sexual and closely related asexual species of *Timema* stick insects. Contrary to theoretical predictions, they found evidence for an overall masculinization of sex-biased gene expression in asexual females, and hypothesized that this is due to shifts in female trait optima levels following the loss of sex. Indeed, for cases where an obligate sexual species loses sexual reproduction, not only does the intensity of intra-locus conflict change, but so does the optimal expression of the new asexual morph, which can differ from that of sexual female. Huylmans et al. (2021) compared gene expression patterns in sexual and asexual species of the brine shrimp *Artemia*. They did not show particular shifts in female- nor male-

biased gene expression after sexuality is lost and explained their results by the fast turnover of sex-biased genes, which had high rates of expression divergence even within sexual *Artemia* species.

A point not addressed in the previous studies (Huylmans et al., 2021; Parker et al., 2019a) is that, in addition to the suppression of sexual conflict, asexuality is associated with a lack of genome recombination. This may affect the evolution of gene expression in several ways. The absence of recombination impedes the efficacy of selection to act upon asexual genomes and prevents the purging of deleterious mutations (Barton & Charlesworth, 1998; Maynard Smith, 1978). Hence, gene expression is likely to evolve toward optimal level at a slower rate for asexual females in asexual lineages. Lack of recombination also leads to independent evolution of asexual lineages that should enhance expression divergence among asexual lineages in the long term. Global gene expression should be more divergent and variable between asexual lineages than between sexual lineages, especially if asexual lineages have diverged long time ago.

In this context, aphids which have evolved partial asexuality long ago and that have then recently lost the sexual phase multiple times are of particular interest. The common sexual ancestor of aphids acquired the ability to alternate between parthenogenetic and sexual reproduction ~250 million years ago (Dixon, 1998; Hales et al., 1997). This alternation of many asexual (parthenogenetic) generations and one sexual generation a year is called cyclical parthenogenesis (CP), and involves three reproductive morphs: viviparous parthenogenetic females, which are present for the main part of the aphid life cycle, males and oviparous sexual females, both of which are produced once a year. Aphids display an XX/X0 sex determination system with X0 males produced by the random elimination of one X chromosome from the germ line (Wilson et al., 1997) and XX sexual and parthenogenetic females that are genetically identical to their parthenogenetic mother. Interestingly, previous work has shown that sexual conflicts may be of importance in CP aphids (Jaquiéry et al., 2013). Mathematical modelling showed the conditions of invasion of sexually antagonistic nonsynonymous mutations favourable to males to be less restrictive on the X than on autosomes (and *vice-versa* for male deleterious alleles), making the X chromosome more favourable for males (Jaquiéry et al., 2013). If the evolution of sex-biased expression can alleviate sexual conflict, the X should be enriched in male-biased genes and this is indeed what is observed (Jaquiéry et al., 2013, 2022; Mathers et al., 2019). The agreement between theoretical predictions and empirical data suggests that sexual conflict is an important player in genome evolution and gene expression in aphids.

Remarkably, some CP aphid lineages have secondarily lost the ability to reproduce sexually, becoming obligate parthenogens (OP). OP lineages produce no sexual females, but parthenogenetic females and sometimes males. Since the egg resulting from sexual reproduction is the sole cold-resistant stage in aphids, only CP lineages can survive in cold-winter regions (Simon et al., 2002, 2010). Contrastingly, OP lineages cannot produce eggs (due to the absence of sexual females), they are thus frost sensitive and are found in mild-winter regions where their ability to maintain parthenogenetic development in winter gives them a demographic advantage over CP lineages (Simon et al., 2002). This geographic separation of CP and OP lineages due to climate has been shown for several aphid species (Blackman, 1974; Dedryver et al., 2001; Defendini et al., 2023; Papura et al., 2003; Simon et al., 1999; Vorburger et al., 2003). As a result, OP males have low opportunity to find mating partners (i.e., sexual females from CP lineages) and to pass on their genes to the next generation. These OP males are supposed to be under relaxed selection, and some of their traits are indeed showing evidence of a slight decay compared to CP males (Defendini et al., 2023). Based on this knowledge, we hypothesized that in OP lineages, only one morph (the parthenogenetic female) is under strong selection, hence conflicts between morphs are relaxed. This may affect the evolution of their transcriptome, in

particular morph-biased genes that may have been involved in an intra-locus conflict that no longer exists in OP lineages. As gene expression in OP parthenogenetic females is not constrained anymore by that in other morphs, we suppose that OP lineages that overexpress parthenogenetic female-biased genes would be favoured by selection. Since the maintenance of differential gene expression between sexes when intra-locus conflict has disappeared might no longer benefit OP lineages, we also expect selection to favour OP lineages whose male gene expression evolves towards parthenogenetic female optimum.

To test our predictions and to investigate more broadly how loss of sex alters gene expression patterns, we performed whole body transcriptome sequencing (RNA-seq) on parthenogenetic females and males in four OP lineages and on parthenogenetic females, males and sexual females in four CP lineages of the pea aphid *Acyrtosiphon pisum*. First, we aimed at evaluating whether the lack of recombination in OP lineages increases gene expression differentiation between OP lineages as genomes of asexual species are expected to evolve independently without mixing since sex loss. For that, we examined the genetic relationships between OP and CP lineages, and tested whether gene expression was more divergent between OP lineages than between CP lineages. Second, to test the hypothesis that the suppression of sexual conflict in OP lineages could result in a change in gene expression patterns, we first identified morph-biased genes using an independent RNA-seq data, and then tested for differences in gene expression between CP and OP parthenogenetic females and CP and OP males for different categories of genes. To better understand the observed changes in gene expression between OP and CP parthenogenetic females or males, we subdivided these morph-biased genes according to their chromosomal location (*X versus autosomes*), the degree of expression bias and the tissue in which they were most expressed.

MATERIALS AND METHODS

Biological samples

Eight genetically distinct lineages of *Acyrtosiphon pisum* were selected from a large panel of lineages used in another study (Defendini et al., 2023), collected from clover fields in France and screened for their ability to produce sexual morphs. Four of them, characterized as OP lineages, were collected in the west of France and the four others collected in the east were characterized as CP lineages. These eight lineages were selected from those that produced enough males to provide the necessary biological material, while being distributed throughout the phylogenetic tree in Defendini et al. (2023).

To produce sexual morphs, the eight lineages were reared under sex-inducing conditions (Le Trionnaire et al., 2009). Briefly, one one-day old larva was placed on a *Vicia faba* plant stored at 18°C with a 16:8 light:dark (L:D) photoperiod regime for one week. Then, the one-week-old individual was moved to short-day conditions (12:12 L:D regime at 18°C). Six days later, four one-day old larvae (i.e., the second generation) were placed on a new plant and maintained under short-day conditions. Once adults, the four individuals produced sexual morphs (males and oviparous sexual females in CP lineages, males only in OP lineages). They were transferred to a new plant every day in order to avoid crowding and to precisely record the age of these individuals. We isolated sexual females from males prior to sexual maturation (at fourth larval stage) to ensure that sexual morphs were virgin. We collected adult individuals two to three days after imaginal moulting. Parthenogenetic females were obtained by rearing the eight lineages under a 16:8 L:D photoperiod regime at 18°C. One one-day-old larva of each lineage was deposited on a *Vicia faba*. Once adult, the female starts to lay offspring. We waited for six offspring to be laid, which takes one day, and the female was then removed from the plant. These individuals were collected two to three days after their imaginal moult. All individuals

were flash-frozen in liquid nitrogen and stored at -80°C. We pooled six to ten males per replicate, three sexual females per replicate and two parthenogenetic females per replicate. Two biological replicates per morph and lineage were generated.

RNA sequencing

We performed a total of 40 RNA extractions (4 CP lineages x 3 morphs x 2 replicates and 4 OP lineages x 2 morphs x 2 replicates) using the SV Total RNA Isolation System (Promega) following manufacturer's instructions. The quality of the RNA samples was checked on Bioanalyzer (Agilent) and quantified on Nanodrop (Thermo Scientific). Libraries were prepared from 500ng of total RNA using the Illumina® Stranded mRNA Prep Ligation library preparation kit, which allows to perform a strand specific sequencing. This protocol includes a first step of polyA selection using magnetic beads. After RNA fragmentation, cDNA synthesis was performed and resulting fragments were used for dA-tailing followed by ligation of RNA Index Anchors. PCR amplification with indexed primers (IDT for Illumina RNA UD Indexes) was finally achieved, to generate the final cDNA libraries. Library quantification and quality assessment was performed using Qubit fluorometric assay (Invitrogen) with dsDNA HS (High Sensitivity) Assay Kit and LabChip GX Touch using a High Sensitivity DNA chip (Perkin Elmer). Libraries were then equimolarly pooled and quantified by qPCR using the KAPA library quantification kit (Roche). Sequencing was carried out on the NovaSeq 6000 instrument from Illumina using paired-end 2 x 100 bp, to obtain around 40 million raw paired-end reads per sample (Suppl. Table 1).

Processing of the RNA-seq data was carried out using the nf-core RNA-seq pipeline version v3.4 (<https://nf-co.re/rnaseq/3.4>). This pipeline was based on Nextflow v21.04.0, using the default settings and adding the option `--remove_ribo_rna`. Reads were mapped to the reference genome *Acyrtosiphon pisum* JIC1 v1.0 (Mathers et al., 2021) available in AphidBase (Legeai et al., 2010) with STAR version 2.6.1d (Dobin et al., 2013). Transcript quantification was performed with featureCounts v1.6.0 (Liao et al., 2014) according to the *A. pisum* JIC1 v1.0 gene annotation with default settings except `-C -p -s 2 -M --fraction`. We then used the R package edgeR version 3.38.4 (Chen et al., 2016) to normalize the libraries (TMM method, $p = 0.75$) and calculate count per million (CPM) on R version 4.3.1 (R Core Team, 2023). Only genes with normalized CPM >1 in at least eight libraries (out of the 40 libraries), i.e., 14,319 genes, were retained in the analyses. In all subsequent analyses, the CPM per gene of the two replicates of the same condition were averaged unless otherwise specified.

Genetic relationships between OP and CP lineages

To gain insight into the genetic relationships between the eight pea aphid lineages, we built a tree using SNP variants retrieved from the transcriptomic data. SNP variants with a genotype quality (GQ) above 20 were identified and further filtered for minimum depth (DP) of 80 in each library with `bcftools` and `vcftools` (Li, 2011). We then built a neighbour-joining tree using Nei distances (Nei, 1978) calculated from 11,449 high-quality SNPs using the `nei.dist` function from `poppr` R package (Kamvar et al., 2014). A total of 1,000 bootstraps were performed using `aboot` function from the same package. One parthenogenetic female library per lineage was used to build the tree (Suppl. Fig. 1A). A Principal Component Analysis (PCA) was also carried out on these eight lineages with the `dudi.pca` function from `ade4` R package (Dray & Dufour, 2007).

Comparison of the divergence in gene expression between CP and OP lineages

The possible effect of the lack of genome recombination on gene expression evolution was tested by comparing transcriptomic distances in CP and in OP lineages. Transcriptomic Euclidean distances

between samples were estimated using the *dist* R function after a normalization step by gene on CPM. We first tested for a correlation between the Nei genetic distance and the male (or female) transcriptomic distance matrices using Mantel tests implemented in the R package *vegan* (Oksanen et al., 2022). We then measured the average distance of a focal lineage to the three other lineages with the same reproductive mode, and tested whether distances differed between reproductive modes using two-sided Mann-Whitney tests. Such tests were performed on genetic distances, transcriptomic distances (measured per sex), and transcriptomic distances corrected for genetic distances (i.e., the ratio of transcriptomic to genetic distances).

Change in morph-biased and unbiased gene expression between OP and CP lineages

To investigate how gene expression may change between OP and CP lineages, we calculated the OP-to-CP log₂-ratio of the mean CPM for each gene between the four parthenogenetic females from OP lineages and the four parthenogenetic females from CP lineages. Genes for which the ratio is greater than zero are expressed more by OP than by CP parthenogenetic females and *vice versa*. To avoid division by zero, we added 0.05 to the denominator and numerator. The same approach was applied to males. We then separated these genes into different classes of expression (using independent RNA-seq data, see the next paragraph), first according to the degree of sex bias of their expression (i.e., morph-biased, morph-limited or unbiased) then as a function of their tissue-biased expression (in head, gonads, legs or not tissue-biased) and finally by their chromosome localization (X *versus* autosomes). For each of these classes of genes, we tested for a deviation of the OP-to-CP log₂ratio from 0 to determine whether some specific classes showed change in expression between OP and CP lineages. For this, we used Wilcoxon signed-rank test with continuity correction and adjustment of *p*-values following Benjamini & Hochberg (1995) implemented in the *wilcox.test* R function.

To identify sex-biased and sex-limited gene classes for the above analyses without relying on our main focal RNA-seq data from the clover host race, we used published RNA-seq data of an *A. pisum* outgroup from the alfalfa host race (Jaquiéry et al., 2013). The dataset consisted of whole-body RNA-seq data with two replicates for each of the three morphs (male, sexual female and parthenogenetic female). RNA-seq data was processed with the same parameters as above and only genes with CPM > 1 in at least two libraries (out of the six libraries) were retained. Following Jaquiéry et al. (2022), a gene was considered as male-biased (or male-limited, respectively) if at least 70% (or 95%, respectively) of its reads were found in the male libraries. We proceeded similarly to identify sexual female- and parthenogenetic female-biased (or limited) genes. Genes that were neither morph-biased nor morph-limited were defined as unbiased (Suppl. Table 2). Reassuringly, most morph-biased genes identified using the *A. pisum* alfalfa host race dataset were also morph-biased in the four CP lineages in the clover host race (Suppl. Fig. 2). Second, to further refine these categories of gene expression with tissue bias information, we used another RNA-seq dataset of the same *A. pisum* alfalfa host race in which three tissues (head, legs and gonads) were investigated in the three morphs (Jaquiéry et al., 2022). Genes with normalized CPM > 1 in at least two libraries (out of the 18 libraries) were retained in the analyses. For each of the morph-biased genes identified with the six whole body libraries, we determined whether or not they also showed tissue-biased expression (gonad-biased, head-biased or leg-biased) in the morph in question. A gene was considered to be tissue-biased in a particular morph if at least 50% of its reads were found in a single tissue of a single morph, i.e., in the two replicate libraries out of the 18 libraries. Finally, we also classified genes as a function of the chromosome type (autosome or X chromosome) on which they were located on the chromosome-scale genome assembly of *A. pisum* (JIC1, Mathers et al., 2021).

Further analyses were conducted without focussing on specific gene categories, identifying differentially expressed (DE) genes between OP and CP lineages separately for males and parthenogenetic females. As these DE analyses presented limited additional findings and mainly serve to verify the outcomes presented above, they are provided solely as supplementary material (Suppl. Text 1). Finally, possible differences in dosage compensation could also contribute to the variation in gene expression between OP and CP lineages. This analysis is presented in Suppl. Text 1.

RESULTS

Genetic relationships between OP and CP lineages

The neighbour-joining tree built with the 11,449 SNPs from the transcriptomic data revealed that the eight OP and CP lineages grouped by reproductive mode (Suppl. Fig. 1A), with good bootstrap support. The first axis of the PCA also separated lineages by their mode of reproduction, which explained 18.5% of the variance (Suppl. Fig. 1B). One OP lineage (OP1) was closer to CP lineages both on the tree and on the PCA. Although mean Nei genetic distances between OP lineages (0.157) were slightly but significantly lower than those measured between CP lineages (0.163, $p = 0.029$, Mann-Whitney test, Suppl. Table 3), we did not observe any shallower branching between OP lineages in the neighbour-joining tree (Suppl. Fig. 1A). These results indicate that it is unlikely for the four OP lineages studied here to have a very recent common ancestor.

No difference in expression divergence between OP and between CP lineages

We evaluated the hypothesis that the absence of recombination would result in higher expression divergence within OP lineages than CP lineages using transcriptomic distances. Parthenogenetic females showed the same average level of divergence regardless of reproductive mode (OP: 138.5, CP: 140.5, $p = 0.88$, Mann-Whitney test, Suppl. Table 3) as males (OP: 137.0, CP: 144.3, $p = 0.20$, Suppl. Table 3). As Nei genetic distances were significantly lower within OP lineages compared to CP lineages (see above), and as transcriptomic distances increased significantly with genetic distances for both parthenogenetic females (Spearman correlation coefficient $r_s = 0.489$, $p = 0.009$, Mantel test) and males ($r_s = 0.479$, $p = 0.013$), transcriptomic distances were corrected for genetic distances. We did not observe differences for parthenogenetic females (OP: 885.4, CP: 862.7, $p = 0.11$, Mann-Whitney test) nor males (OP: 877.0, CP: 885.9, $p = 0.89$).

OP parthenogenetic females overexpress sexual female-biased genes

OP parthenogenetic females expressed sexual female-biased genes at a higher level than CP parthenogenetic females (Fig. 1), with a median OP-to-CP log₂-ratio (hereafter referred to as *Mdn*) of 0.121 (Wilcoxon signed-rank test against 0, $p < 10^{-15}$). Remarkably, this pattern was consistent for all OP lineages when pairwise comparisons were made (Suppl. Fig. 3). In contrast, no substantial shift was observed for the other classes of genes (i.e., male-biased, parthenogenetic female-biased and unbiased genes) though the effect was significant in each case due to large sample size (Fig. 1 and 2A). The unbiased genes differed from the morph-biased genes by their much lower variance, indicating higher stability of expression.

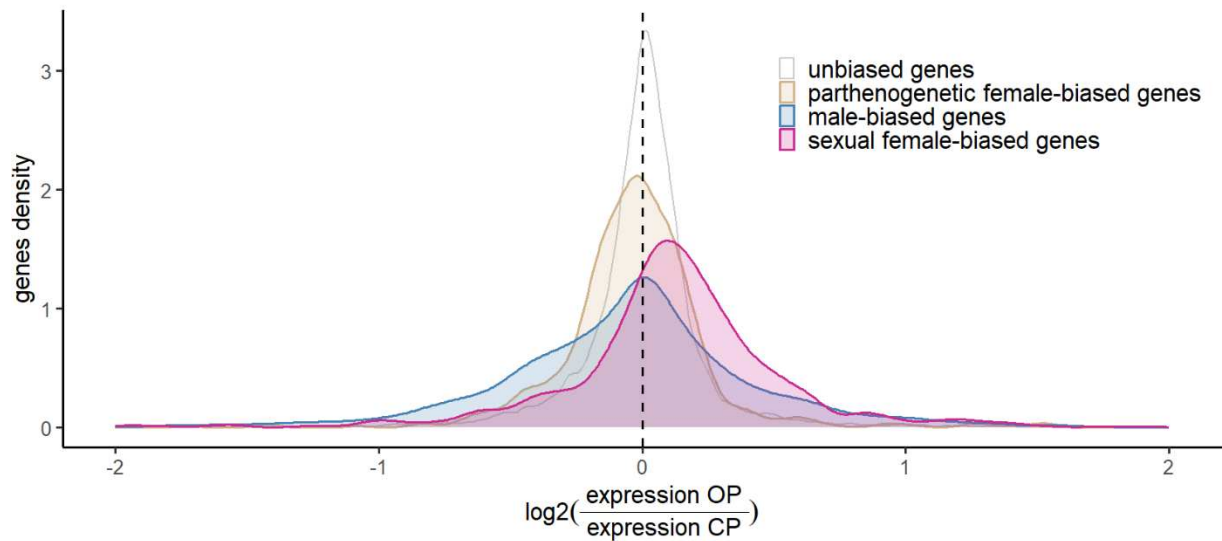


Figure 1. Gene expression changes between parthenogenetic females from OP and CP lineages of the pea aphid for different categories of genes (in different colours). The OP-to-CP log₂-ratio of gene expression (in CPM) is shown on the x-axis. Colours discriminate between sexual female-biased genes (pink), parthenogenetic female-biased genes (beige), male-biased genes (blue) and unbiased genes (white).

To test whether a specific category of genes was responsible for the unexpected increased expression of sexual female-biased genes in OP parthenogenetic females, we first separated these genes by the degree of their morph bias: sexual female-biased genes (i.e. F+; $Mdn = 0.126$, $p < 10^{-15}$) but not sexual female-limited genes (i.e. F++; $Mdn = 0.046$, $p > 0.6$) explained the high OP-to-CP expression ratio (Fig. 2B). We then looked for an effect of the chromosome type: both sexual female-biased genes assigned to autosomes ($Mdn = 0.114$, $p < 10^{-9}$) and to the X chromosome ($Mdn = 0.128$, $p < 10^{-6}$) showed a high OP-to-CP expression ratio, with no significant difference between the two types of chromosomes (Wilcoxon rank sum test, $p = 0.58$, Fig. 2C). Finally, the sub-classification of sexual female-biased genes according to their expression in different tissues showed that genes more expressed in sexual female ovaries contributed the most to the high OP-to-CP expression ratio ($Mdn = 0.138$, $p < 10^{-15}$, Fig. 2D). For comparison, the same type of analyses was performed for male-biased and parthenogenetic female-biased genes, but no strong trend was found (see Suppl. Fig. 4). In summary, parthenogenetic females from OP lineages unexpectedly overexpress sexual female-biased genes that are mainly expressed in female ovaries.

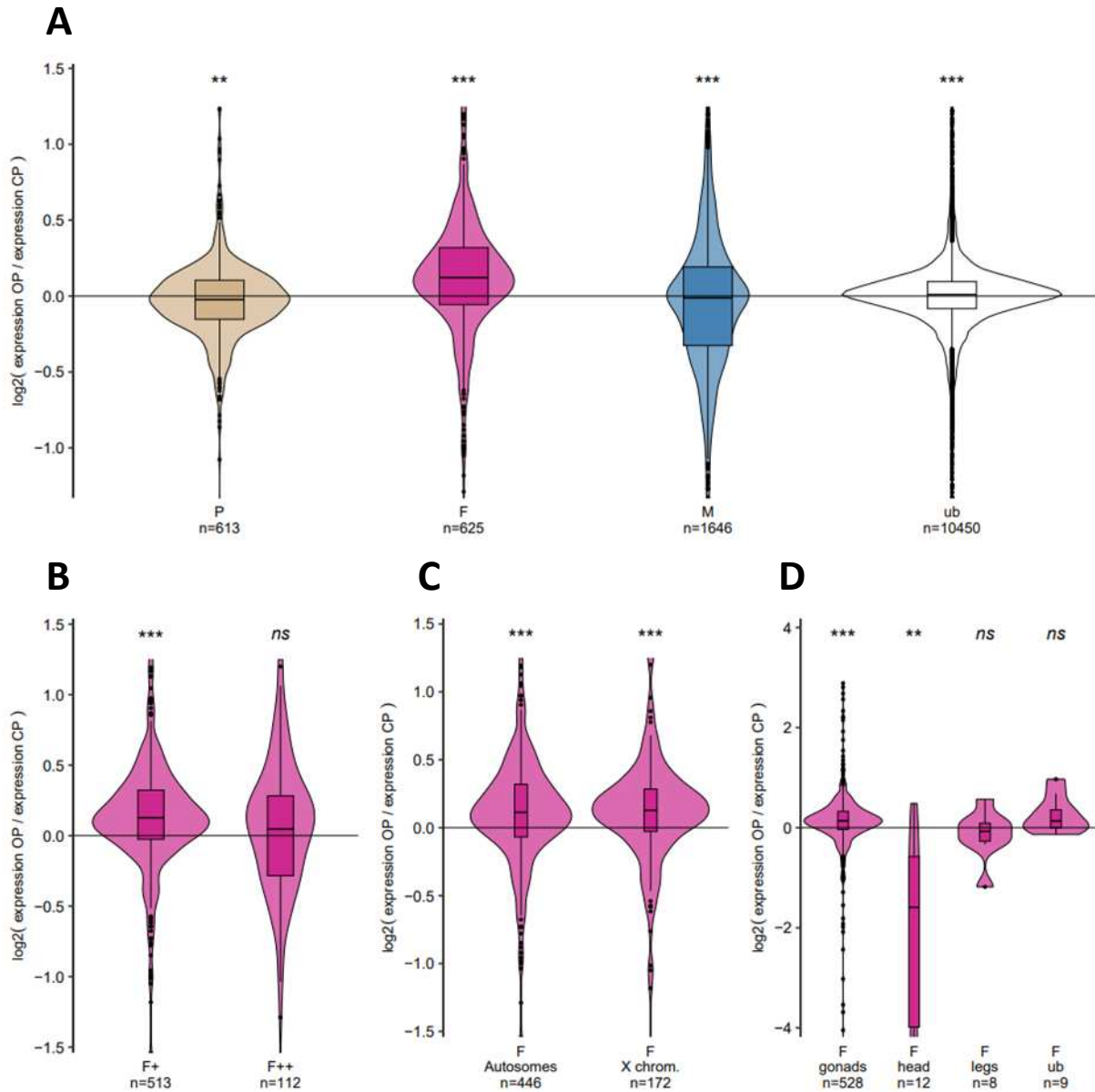


Figure 2. Log₂-ratio of OP-to-CP gene expression in parthenogenetic females of the pea aphid for different categories of genes. OP-to-CP expression ratio as a function of: **(A)** morph-biased status of genes, **(B)** degree of morph-bias (+: biased only or ++: limited) **(C)** chromosomal location of genes, and **(D)** morph- and tissue-bias status of genes. Colours discriminate between sexual female-biased genes (F, in pink), parthenogenetic female-biased genes (P, in beige), male-biased genes (M, in blue) and unbiased genes (ub, in white). The significance level of Wilcoxon signed-rank tests against 0 is indicated by asterisks (***) <math>p < 0.001</math>, (**) <math>p < 0.01</math>, (ns) $p \geq 0.05$).

OP males overexpress parthenogenetic female-biased but also male-biased genes

Our analyses revealed that OP males over-expressed parthenogenetic female-biased genes in comparison to CP males ($Mdn = 0.132$, $p < 10^{-6}$, Fig. 3). Moreover, despite the low median OP-to-CP expression ratio for male-biased genes ($Mdn = -0.004$, $p = 0.006$), we observed a bimodal density

distribution of male-biased gene expression, with a secondary peak for expression ratio around ~ 1.3 (Fig. 3) that corresponded to a subset of male-biased genes overexpressed by OP males. No substantial shift was observed for sexual female-biased and unbiased genes though the effect was always significant due to large sample size (Fig. 3 and 4A). The unbiased genes were again distinguished from the morph-biased genes by their much lower variance.

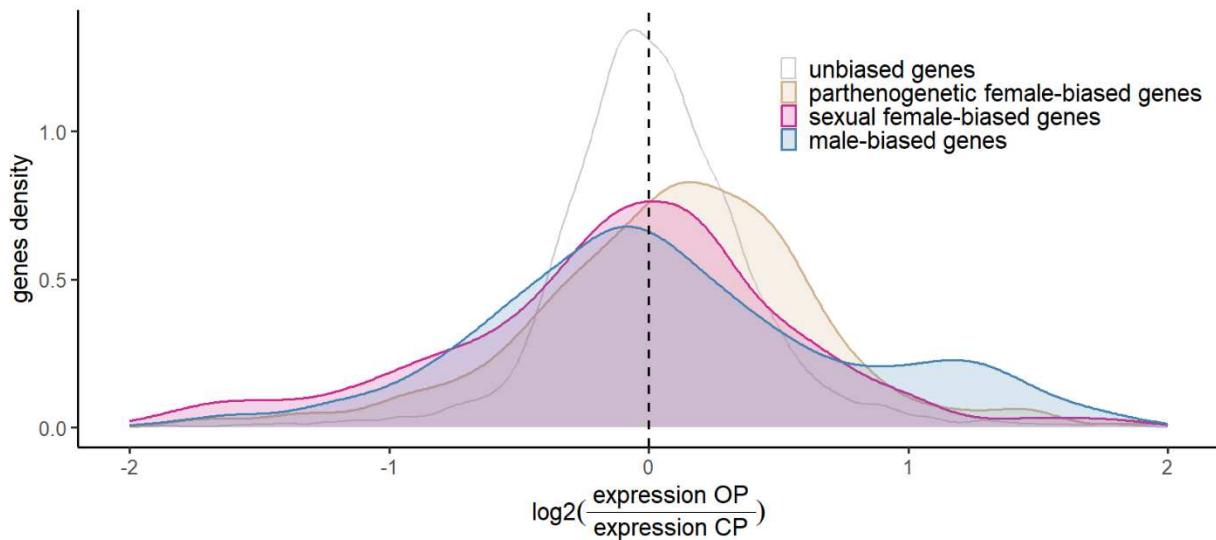


Figure 3. Gene expression changes between males from OP and CP lineages of the pea aphid for different categories of genes (in different colours). The OP-to-CP log₂-ratio of gene expression (in CPM) is shown on the x-axis. Colours discriminate between sexual female-biased genes (pink), parthenogenetic female-biased genes (beige), male-biased genes (blue) and unbiased genes (white).

We next wanted to understand which gene classes contributed to the increased expression of parthenogenetic female-biased genes and the increased expression of a subset of male-biased genes in OP males. Interestingly, the increased expression of parthenogenetic female-biased genes by OP males was common to most of the OP-CP pairwise comparisons (Suppl. Fig. 5), demonstrating the consistency of this pattern. Both parthenogenetic female-biased only genes ($P+$, $Mdn = 0.133$, $p < 10^{-5}$) and parthenogenetic female-limited genes ($P++$, $Mdn = 0.129$, $p = 0.026$) were significantly more expressed by OP males (Fig. 4B). Contrastingly, only autosomal parthenogenetic female-biased genes were more expressed by OP males ($Mdn = 0.145$, $p < 10^{-6}$), with X-linked genes showing no effect ($p = 0.28$, Fig. 4C). Finally, the sub-classification of parthenogenetic female-biased genes according to their expression in different tissues showed that genes mostly expressed in parthenogenetic female ovaries ($Mdn = 0.136$, $p = 0.02$) as well as those expressed similarly in all three tissues ($Mdn = 0.095$, $p = 0.004$) contributed to the high OP-to-CP expression ratio in males (power was low for the two other classes of genes, Fig. 4D).

Our analyses also revealed that the secondary peak observed for the class of male-biased genes was mostly due to male-limited genes. Indeed, OP males slightly underexpressed male-biased only genes ($M+$, $Mdn = -0.114$, $p < 10^{-8}$) but clearly overexpressed male-limited genes ($M++$, $Mdn = 0.379$, $p < 10^{-15}$) relative to CP males (Fig. 4E). Male-biased genes assigned to the X chromosome were overexpressed by OP males ($Mdn = 0.106$, $p < 10^{-7}$) while autosomal male-biased genes were not ($p = 0.19$, Fig. 4F). Among all male-biased genes with tissue bias, those that were testis-biased were more

expressed by OP males than CP males ($Mdn = 0.817$, $p < 10^{-15}$, Fig. 4G). For comparison, the same type of analysis (separating genes by chromosome, tissue or expression type) was carried out for sexual female-biased genes, but no strong trend was found (see Suppl. Fig. 6).

In summary, OP males overexpressed parthenogenetic female-biased genes, as expected under the relaxation of sexual conflict following sex loss. Surprisingly, OP males overexpressed genes that are more specifically expressed in the testes (which additionally tend to be male-limited and more frequent on the X chromosome, Jaquiéry et al., 2022).

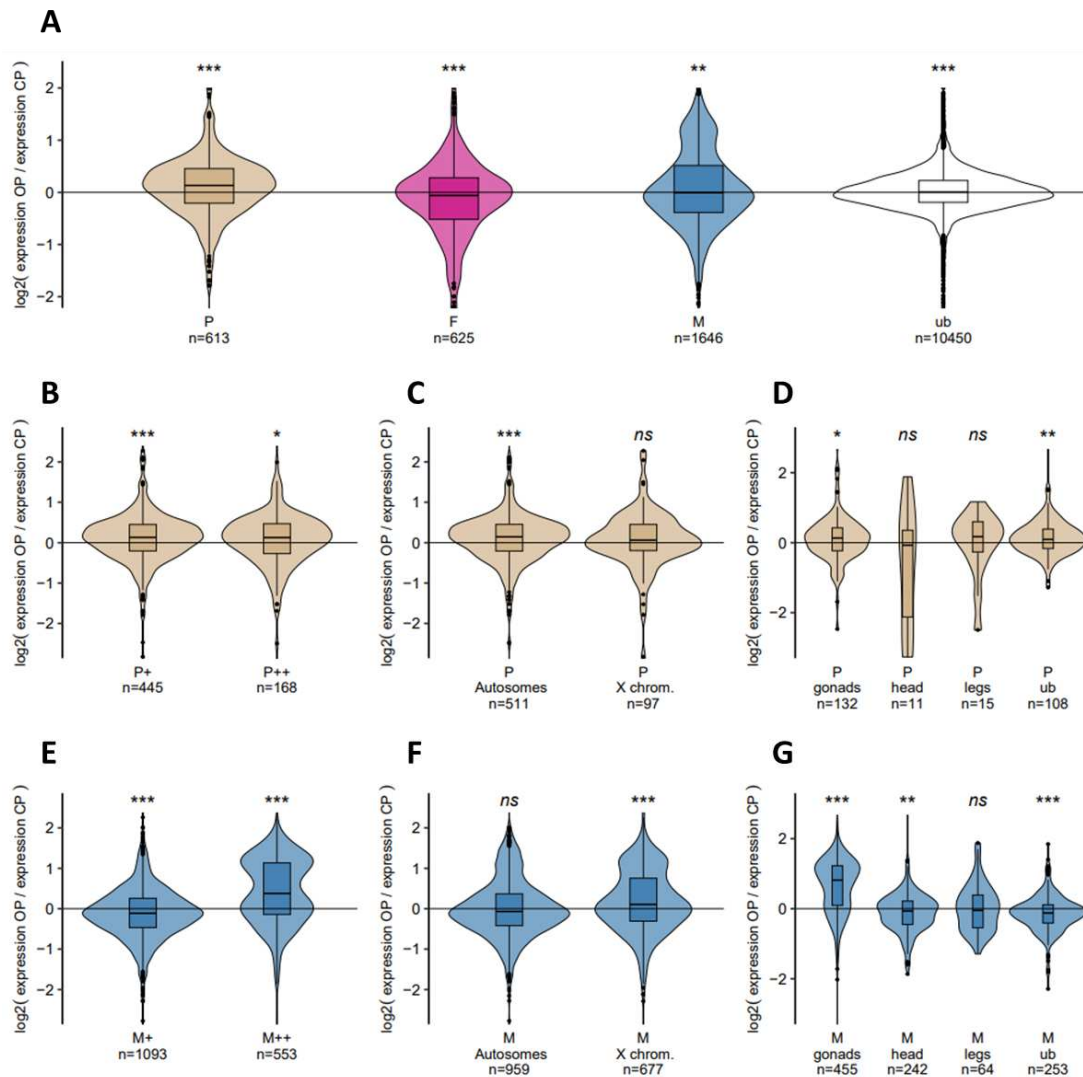


Figure 4. Log₂-ratio of OP-to-CP gene expression in males of the pea aphid for different categories of genes. OP-to-CP expression ratio as a function of: **(A)** morph-biased status of genes, **(B, E)** degree of morph-bias (+: biased only or ++: limited) **(C, F)** chromosomal location of genes, and **(D, G)** morph- and tissue-bias status of genes. Colours discriminate between sexual female-biased genes (F, in pink), parthenogenetic female-biased genes (P, in beige), male-biased genes (M, in blue) and unbiased genes (ub, in white). The significance level of Wilcoxon signed-rank tests against 0 is indicated by asterisks (*** < 0.001, ** < 0.01, * < 0.05, ns ≥ 0.05).

DISCUSSION

Here, we investigated how the relaxation of sexual conflict in lineages that have lost sex might affect the evolution of gene expression at the whole-genome scale, by comparing gene expression patterns in sexual (CP) and asexual (OP) lineages of the pea aphid. We predicted a shift in gene expression towards the parthenogenetic female optimum in asexual (OP) lineages. Consistent with this prediction, we found that OP males tended to express genes normally overexpressed in parthenogenetic females at higher levels than CP males. Unexpectedly, OP males also overexpressed testis genes, and OP parthenogenetic females overexpressed genes that are normally overrepresented in sexual female ovaries. In addition to reduced sexual conflict, the absence of recombination in OP genomes should result in an independent evolution of lineages, leading to a greater divergence of expression among OP relative to CP lineages. However, we did not observe such an effect.

Morph-biased but not unbiased genes are affected by the transition to permanent asexuality

Analyses of the shift in gene expression associated with the transition to permanent asexuality revealed contrasting results for the different classes of genes (i.e., unbiased and morph-biased) defined according to their expression pattern in a related lineage of the pea aphid complex. Genes with unbiased expression showed little change in expression between OP and CP lineages, as indicated by an OP-to-CP expression ratio close to 0 and its low variance. This class was also underrepresented among the genes identified as differentially expressed between OP and CP morphs (Suppl. Text 1). The unbiased gene class is likely to contain a high proportion of essential genes or genes with pleiotropic effects that are therefore under strong stabilising selection for optimal expression and would be little affected by a change in the selective regime, i.e., the disappearance of sexual conflict (Mank et al., 2008; Mank & Ellegren, 2009). In contrast, several of the classes of genes defined as morph-biased (either male-, sexual female- or parthenogenetic female-biased) shifted expression between reproductive modes, and some of these classes were also overrepresented in DE genes (see Suppl. Text 1). Sex-biased genes are known to be labile in their expression (Ellegren & Parsch, 2007). They are often thought to have functions related to sexual dimorphism and reproduction, and may be involved in past or ongoing intra- or inter-locus sexual conflicts. Therefore, the ongoing arm race between the sexes could explain the high rate of change in expression or sequence generally observed (Grath & Parsch, 2016; Whittle & Extavour, 2019; Wright et al., 2018). These genes would be anticipated to be particularly responsive to a change in the selective regime, namely the disappearance of sexual conflict following sex loss, as we observed here. However, another non-exclusive hypothesis is that sex-biased genes correspond to genes that were once dispensable (their dispensability allowed them to acquire sex-biased expression), so that their expression (and sequences) could still evolve rapidly at low fitness cost (Mank & Ellegren, 2009).

OP males overexpress genes normally overexpressed in parthenogenetic females

OP males overexpressed parthenogenetic female-biased genes compared to CP males, as shown by the OP-to-CP gene expression ratio, a trend consistent regardless of the chromosome type, the type of expression (biased or limited) or the tissue involved. This is consistent with the prediction of a feminization of male transcriptome under asexuality, with optimal fitness for parthenogenetic females favoured over males, who are no longer under selection. Such feminization of the male transcriptome has also been observed following relaxation of selective pressure on males, resulting from an artificial transition from polygyny to monogamy in *D. melanogaster* (Hollis et al., 2014). Interestingly, OP pea aphid males also showed a slight decrease in reproductive performances compared to CP males, suggesting relaxed selection on male traits (Defendini et al., 2023). These results thus show that the

effects of relaxed selection on OP males (including the ones resulting from relaxed sexual conflict) can be detected across various phenotypic dimensions.

Asexual morphs overexpress genes primarily expressed in the gonads of sexual morphs

Surprisingly, males from OP lineages are expressing testis genes more than CP males, which goes against our expectations (see also Dean et al., 2018). This is unexpected because OP males are thought to have evolved in a context of relaxed selection. Similarly, OP parthenogenetic females unexpectedly overexpressed sexual female-biased genes normally expressed in sexual female ovaries compared to CP parthenogenetic females. Sexual female-biased genes were also overrepresented among DE genes between OP and CP parthenogenetic females (see Suppl. Text 1). Remarkably, this trend of gonadal gene overexpression by asexual lineages was observed across all lineages, indicating that it is a general pattern.

Most studies (including our own) that have examined how changes in selection regimes during reproduction affect the expression of sex-specific genes have revealed unexpected results. In asexual *Timema* species, parthenogenetic females underexpress sexual female-biased genes but overexpress male-biased genes (Parker et al., 2019a), whereas no change occurs in asexual *Artemia* (Huylmans et al., 2021). Unexpectedly, enforced monogamy in the polygynous *D. pseudoobscura* tends to masculinize the female transcriptome (Veltsos et al., 2017). Only the work on *D. melanogaster* fits the predictions, with a feminization of male and female transcriptomes under enforced monogamy (Hollis et al., 2014). Various hypotheses have been put forward to explain these unexpected patterns. In particular, in *Timema* and *Artemia*, the loss of sex is confounded with the emergence of a new reproductive morph, the parthenogenetic female morph. The fitness optima for this newly emerged morph should not necessarily be similar to that of sexual female (the former no longer needs to mate and is able to produce diploid eggs through modified cellular processes), which could raise different expectations. Parker et al. (2019a) therefore argued that the observed masculinization of the transcriptome of asexual females is not directly due to the release of sexual conflict in asexual lineages, but rather to a change in the optimal female phenotype. However, in the pea aphid, the difference in optima for parthenogenetic females in OP and CP lineages is unlikely to explain changes in expression because parthenogenetic females are originally part of the CP life cycle.

Several hypotheses, whether adaptive or not, can be put forward to explain the overexpression of testis genes in OP males and sexual female ovary genes in OP parthenogenetic females. A first non-adaptive hypothesis is that these expression profiles in OP lineages may have been inherited from a common (and relatively recent) OP ancestor that, by chance, already had a slightly atypical expression profile. This would explain the consistency of the overexpression of sexual female ovary genes (in parthenogenetic females) and testis genes (in males) across all four OP lineages. Population history can indeed be a major source of variation in gene expression, as shown in asexual and sexual populations of freshwater snails (McElroy et al., 2022). However, our analyses of the genetic relationships between the four OP and four CP lineages based on >10,000 SNPs (current study) and a larger panel of 23 OP and 76 CP lineages using a lower number of markers (15 microsatellite loci, Defendini et al., 2023) contradict this hypothesis. Although the SNP tree grouped the four OP lineages with excellent bootstrap support, one OP lineage (OP1) was the closest to the CP lineages. On the 15-microsatellite tree, the OP1 lineage is closer to CP lineages than to OP lineages, suggesting possible independent transitions to permanent asexuality. Furthermore, although the genetic distances among the OP lineages are marginally lower than those of the CP lineages, the values are quite similar, suggesting that the OP lineages have been evolving for a comparable duration as the CP lineages. We

can therefore rule out the possible effect of recent common ancestry as an explanation for the peculiar expression of sexual gonad genes in OP lineages. A second non-adaptive hypothesis for the shift in expression of genes in the sexual gonads could be that the dosage compensation system has been disrupted in OP lineages. Indeed, dosage compensation would be expected to decay as the lack of reproductive opportunities for OP males leads to relaxed selection on males. However, we found no major change in dosage compensation in OP compared to CP lineages, as assessed by comparing the male-to-female expression ratio for autosomal and X-linked genes between OP and CP lineages (see Suppl. Text 1). A third hypothesis could be that in OP lineages, the genetic program that normally controls gonadal genes in CP sexual morphs is dysregulated. Indeed, the absence of sexual females in OP lineages and the presumed lack of reproductive opportunity for OP males means that selection is no longer acting on these morphs, and in particular on the genes that are specific to these morphs. Small changes in one or more upstream regulatory genes may be sufficient to affect a large number of downstream genes. Evolutionary changes in regulatory regions are frequent (Johnson, 2017), with direct evidence found in animals (Erkenbrack & Davidson, 2015), fungi (Wilinski et al., 2017) and plants (Muiño et al., 2016). Alternatively, sex-biased genes, especially those restricted to gonads, evolve rapidly at the protein level, as observed in many clades (Ellegren & Parsch, 2007; Grath & Parsch, 2012; Mank et al., 2007; Meisel, 2011; Whittle & Johannesson, 2013). Hence, genes initially linked to sexual reproduction due to their predominant expression in sexual morph gonads within CP lineages, may have evolved functional modifications that confer greater advantages to parthenogenetic females. A detailed study of these sexual female ovary genes is required to elucidate their precise functions, yet the substantial number of genes with unknown function in the pea aphid poses a significant challenge. Finally, the shift in sexual morph-biased genes in OP morphs may be inherent to the transition to obligate parthenogenesis. The path to asexuality may be constrained in this species, requiring certain variants to persist without sex, and if these variants have large effects on the regulation of gonad gene networks, this could explain the differences in expression observed between OP and CP morphs. In *Timema* stick insects, repeated and independent events of sex loss are associated with convergent gene expression changes, indicating that the evolutionary pathway to asexuality is highly constrained and requires changes in the same key genes (Parker et al., 2019b).

Non-recombining OP lineages do not show increased divergence in gene expression

Loss of sex causes the genome to stop recombining, allowing OP lineages to evolve independently, which should increase expression divergence between them. The lack of recombination also reduces the efficiency of selection, resulting in deleterious mutations accumulation with potential snowballing effects on downstream genes. Few studies have examined the evolution of gene expression within populations that differ in recombination rate. Theoretical models predict that lack of recombination would eventually lead to haploidization of expression, with one of the two alleles at a diploid locus contributing the majority of expression via an enhancer runaway divergence process (Fyon & Lenormand, 2018). However, no prediction was made regarding the expected expression divergence between asexual lineages. Empirical evidence for gene expression divergence in apomictic (asexual) plant lineages is provided by Ferreira de Carvalho et al. (2016), but no comparisons were made with sexual counterparts. Ma et al. (2020) investigated the allele-specific expression in non-recombining regions of the sex chromosomes of a fungus and showed evidence for a relationship between expression divergence between alleles and allele degeneration. To our knowledge, our study is the first to compare gene expression divergence within different reproductive modes. Contrary to our expectations of increased divergence between OP lineages, we observed no effect of the reproductive

mode on transcriptomic distances, corrected or not for genetic distances. Mathematical modelling may be necessary to verify the accuracy of our predictions regarding the evolution of gene expression in the absence of recombination. Purifying selection may also play a role, wherein OP lineages exhibiting lesser expression divergence experience enhanced fitness, consequently limiting the divergence among OP lineages. In any case, it would be necessary to work on a larger panel of lineages, as four lineages per reproductive mode is a low number for estimating and comparing transcriptomic divergence.

Conclusion

Our study highlights the evolutionary implications of the loss of sexual reproduction, which results in the cessation of intra-locus sexual conflict and recombination. We found substantial changes in the expression of morph-biased genes, yet not always in the expected direction. This suggests that processes other than the reduction of sexual conflict may influence the evolution of gene expression in asexual lineages, and these need to be investigated further. It would also be interesting to investigate whether an 'enhancer runaway divergence' process occurs in OP lineages due to the accumulation of deleterious mutations and the lack of recombination, and whether this process contributes to the expression changes. In addition, further investigation at the DNA sequence level would be valuable, for example, to test whether genes typically restricted to sexual female ovaries accumulate deleterious mutations in OP lineages due to the lack of selection, or conversely, exhibit signs of positive selection, thus unveiling adaptive changes.

DATA ACCESSIBILITY

Raw reads have been deposited in the SRA, under the BioProject accession number PRJNA1024923 and will be available after publication. The processed data (read counts and genotype data) and scripts will soon be available on Zenodo and the remaining data is included as Supplementary Material.

ACKNOWLEDGMENTS AND FUNDING STATEMENT

This work was supported by grants from the French National Research Agency (ANR) SexAphid (ANR-09-GENM-017-001), Mecadapt (ANR-11-BSV7-005-01), the INRAE-SPE Department (AAP CyclEvol and half a PhD grant for HD) and the Region Bretagne (half a PhD grant for HD). High-throughput sequencing was performed by the ICGex NGS platform of the Institut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the French National Research Agency ("Investissements d'Avenir" program), by the ITMO-Cancer Aviesan (Plan Cancer III) and by the SiRIC-Curie program (SiRIC Grant INCa-DGOS-465 and INCa-DGOS-Inserm_12554). Data management, quality control and primary analysis were performed by the Bioinformatics platform of the Institut Curie. The authors declare no conflicting interest.

AUTHOR CONTRIBUTION

JJ and JCS conceived and designed the study. HD, JJ, NPL, SL, SB generated the data. HD and SR analysed the data. HD, JJ, and JSC drafted the manuscript. All authors reviewed and revised the manuscript.

REFERENCES

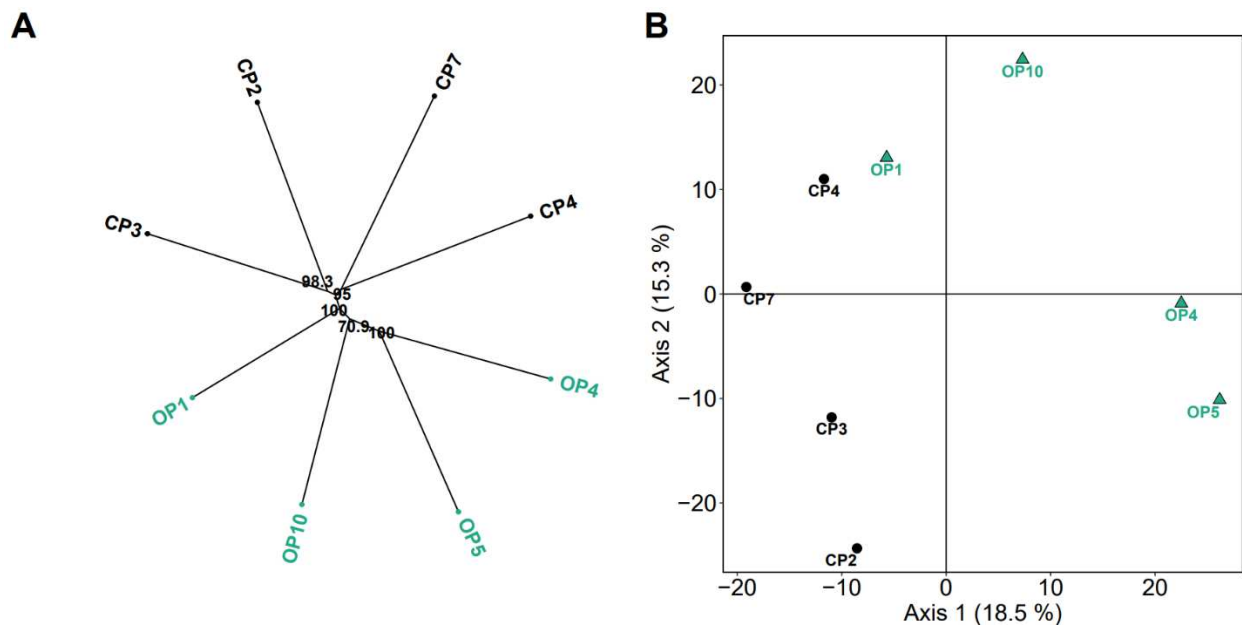
- Barton, N. H., & Charlesworth, B. (1998). Why sex and recombination? *Science*, *281*(5385), 1986–1990. <https://doi.org/10.1126/science.281.5385.1986>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Blackman, R. L. (1974). Life-cycle variation of *Myzus persicae* (Sulz.) (*Hom.*, *Aphididae*) in different parts of the world, in relation to genotype and environment. *Bulletin of Entomological Research*, *63*(4), 595–607. <https://doi.org/10.1017/S0007485300047830>
- Bonduriansky, R., & Chenoweth, S. F. (2009). Intralocus sexual conflict. *Trends in Ecology & Evolution*, *24*(5), 280–288. <https://doi.org/10.1016/j.tree.2008.12.005>
- Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, *5*, 1438. <https://doi.org/10.12688/f1000research.8987.2>
- Cheng, C., & Kirkpatrick, M. (2016). Sex-specific selection and sex-biased gene expression in humans and flies. *PLOS Genetics*, *12*(9), e1006170. <https://doi.org/10.1371/journal.pgen.1006170>
- Chippindale, A. K., Gibson, J. R., & Rice, W. R. (2001). Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proceedings of the National Academy of Sciences*, *98*(4), 1671–1675. <https://doi.org/10.1073/pnas.98.4.1671>
- Dean, R., Hammer, C., Higham, V., & Dowling, D. K. (2018). Masculinization of gene expression is associated with male quality in *Drosophila melanogaster*. *Evolution*, *72*(12), 2736–2748. <https://doi.org/10.1111/evo.13618>
- Dedryver, C.-A., Hullé, M., Le Gallic, J.-F., Caillaud, M. C., & Simon, J.-C. (2001). Coexistence in space and time of sexual and asexual populations of the cereal aphid *Sitobion avenae*. *Oecologia*, *128*(3), 379–388. <https://doi.org/10.1007/s004420100674>
- Defendini, H., Rimbault, M., Mahéo, F., Cloteau, R., Denis, G., Mieuze, L., Outreman, Y., Simon, J.-C., & Jaquiéry, J. (2023). Evolutionary consequences of loss of sexual reproduction on male-related traits in parthenogenetic lineages of the pea aphid. *Molecular Ecology*, *32*(13), 3672–3685. <https://doi.org/10.1111/mec.16961>
- Dixon, A. F. G. (1998). *Aphid Ecology* (2nd ed.). Chapman & Hall.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- 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>
- Ellegren, H., & Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, *8*(9), 9. <https://doi.org/10.1038/nrg2167>
- Erkenbrack, E. M., & Davidson, E. H. (2015). Evolutionary rewiring of gene regulatory network linkages at divergence of the echinoid subclasses. *Proceedings of the National Academy of Sciences*, *112*(30), E4075–E4084. <https://doi.org/10.1073/pnas.1509845112>
- Ferreira de Carvalho, J., Oplaat, C., Pappas, N., Derks, M., de Ridder, D., & Verhoeven, K. J. F. (2016). Heritable gene expression differences between apomictic clone members in *Taraxacum officinale*: Insights into early stages of evolutionary divergence in asexual plants. *BMC Genomics*, *17*(1), 203. <https://doi.org/10.1186/s12864-016-2524-6>
- Fyon, F., & Lenormand, T. (2018). Cis-regulator runaway and divergence in asexuals. *Evolution*, *72*(3), 426–439. <https://doi.org/10.1111/evo.13424>
- Grath, S., & Parsch, J. (2012). Rate of amino acid substitution is influenced by the degree and conservation of male-biased transcription over 50 Myr of *Drosophila* evolution. *Genome Biology and Evolution*, *4*(3), 346–359. <https://doi.org/10.1093/gbe/evs012>
- Grath, S., & Parsch, J. (2016). Sex-biased gene expression. *Annual Review of Genetics*, *50*(1), 29–44. <https://doi.org/10.1146/annurev-genet-120215-035429>

- Griffin, R. M., Dean, R., Grace, J. L., Rydén, P., & Friberg, U. (2013). The Shared Genome Is a Pervasive Constraint on the Evolution of Sex-Biased Gene Expression. *Molecular Biology and Evolution*, *30*(9), 2168–2176. <https://doi.org/10.1093/molbev/mst121>
- Hales, D. F., Tomiuk, J., Woehrmann, K., & Sunnucks, P. (1997). Evolutionary and genetic aspects of aphid biology: A review. *EJE*, *94*(1), 1–55.
- Holland, B., & Rice, W. R. (1999). Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. *Proceedings of the National Academy of Sciences*, *96*(9), 5083–5088. <https://doi.org/10.1073/pnas.96.9.5083>
- Hollis, B., Houle, D., Yan, Z., Kawecki, T. J., & Keller, L. (2014). Evolution under monogamy feminizes gene expression in *Drosophila melanogaster*. *Nature Communications*, *5*(1), 1. <https://doi.org/10.1038/ncomms4482>
- Huylmans, A. K., Macon, A., Hontoria, F., & Vicoso, B. (2021). Transitions to asexuality and evolution of gene expression in *Artemia* brine shrimp. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1959), 20211720. <https://doi.org/10.1098/rspb.2021.1720>
- Innocenti, P., & Morrow, E. H. (2010). The sexually antagonistic genes of *Drosophila melanogaster*. *PLOS Biology*, *8*(3), e1000335. <https://doi.org/10.1371/journal.pbio.1000335>
- Jaquiéry, J., Rispe, C., Roze, D., Legeai, F., Trionnaire, G. L., Stoeckel, S., Mieuzet, L., Silva, C. D., Poulain, J., Prunier-Leterme, N., Séguens, B., Tagu, D., & Simon, J.-C. (2013). Masculinization of the X chromosome in the pea aphid. *PLOS Genetics*, *9*(8), e1003690. <https://doi.org/10.1371/journal.pgen.1003690>
- Jaquiéry, J., Simon, J.-C., Robin, S., Richard, G., Peccoud, J., Boulain, H., Legeai, F., Tanguy, S., Prunier-Leterme, N., & Letrionnaire, G. (2022). Masculinization of the X-chromosome in aphid soma and gonads. *Peer Community Journal*, *2*. <https://doi.org/10.24072/pcjournal.166>
- Johnson, A. D. (2017). The rewiring of transcription circuits in evolution. *Current Opinion in Genetics & Development*, *47*, 121–127. <https://doi.org/10.1016/j.gde.2017.09.004>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. <https://doi.org/10.7717/peerj.281>
- Lande, R. (1980). Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution*, *34*(2), 292–305. <https://doi.org/10.2307/2407393>
- Le Trionnaire, G., Francis, F., Jaubert-Possamai, S., Bonhomme, J., De Pauw, E., Gauthier, J.-P., Haubruge, E., Legeai, F., Prunier-Leterme, N., Simon, J.-C., Tanguy, S., & Tagu, D. (2009). Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid. *BMC Genomics*, *10*(1), 456. <https://doi.org/10.1186/1471-2164-10-456>
- Legeai, F., Shigenobu, S., Gauthier, J.-P., Colbourne, J., Rispe, C., Collin, O., Richards, S., Wilson, A. C., Murphy, T., & Tagu, D. (2010). AphidBase: A centralized bioinformatic resource for annotation of the pea aphid genome. *Insect Molecular Biology*, *19*(s2), 5–12. <https://doi.org/10.1111/j.1365-2583.2009.00930.x>
- 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>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Ma, W.-J., Carpentier, F., Giraud, T., & Hood, M. E. (2020). Differential gene expression between fungal mating types is associated with sequence degeneration. *Genome Biology and Evolution*, *12*(4), 243–258. <https://doi.org/10.1093/gbe/evaa028>
- Mank, J. E., & Ellegren, H. (2009). Are sex-biased genes more dispensable? *Biology Letters*, *5*(3), 409–412. <https://doi.org/10.1098/rsbl.2008.0732>
- Mank, J. E., Hultin-Rosenberg, L., Axelsson, E., & Ellegren, H. (2007). Rapid evolution of female-biased, but not male-biased, genes expressed in the avian brain. *Molecular Biology and Evolution*, *24*(12), 2698–2706. <https://doi.org/10.1093/molbev/msm208>

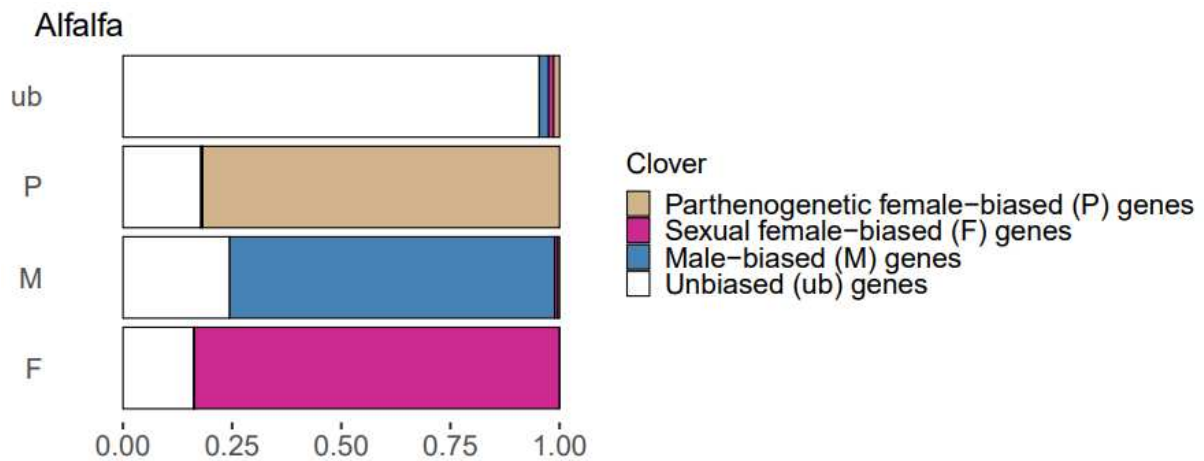
- Mank, J. E., Hultin-Rosenberg, L., Zwahlen, M., & Ellegren, H. (2008). Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *The American Naturalist*, *171*(1), 1. <https://doi.org/10.1086/523954>
- Mathers, T. C., Mugford, S. T., Percival-Alwyn, L., Chen, Y., Kaithakottil, G., Swarbreck, D., Hogenhout, S. A., & van Oosterhout, C. (2019). Sex-specific changes in the aphid DNA methylation landscape. *Molecular Ecology*, *28*(18), 4228–4241. <https://doi.org/10.1111/mec.15216>
- Mathers, T. C., Wouters, R. H. M., 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>
- Maynard Smith, J. (1978). *The Evolution of Sex*. Cambridge University Press.
- McElroy, K. E., Bankers, L., Soper, D., Hehman, G., Boore, J. L., Logsdon, J. M., & Neiman, M. (2022). Patterns of gene expression in ovaries of sexual vs. asexual lineages of a freshwater snail. *Frontiers in Ecology and Evolution*, *10*. <https://www.frontiersin.org/articles/10.3389/fevo.2022.845640>
- Meisel, R. P. (2011). Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. *Molecular Biology and Evolution*, *28*(6), 1893–1900. <https://doi.org/10.1093/molbev/msr010>
- Muiño, J. M., de Bruijn, S., Pajoro, A., Geuten, K., Vingron, M., Angenent, G. C., & Kaufmann, K. (2016). Evolution of DNA-binding sites of a floral master regulatory transcription factor. *Molecular Biology and Evolution*, *33*(1), 185–200. <https://doi.org/10.1093/molbev/msv210>
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, *89*(3), 583–590. <https://doi.org/10.1093/genetics/89.3.583>
- Oksanen, J., Simpson, G., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P., Hara, R., Solymos, P., STEVENS, H., Szöcs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Cáceres, M., Durand, S., & Weedon, J. (2022). *Vegan community ecology package version 2.6-2*.
- Papura, D., Simon, J.-C., Halkett, F., Delmotte, F., Le Gallic, J.-F., & Dedryver, C.-A. (2003). Predominance of sexual reproduction in Romanian populations of the aphid *Sitobion avenae* inferred from phenotypic and genetic structure. *Heredity*, *90*, 5. <https://doi.org/10.1038/sj.hdy.6800262>
- Parker, D. J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M., & Schwander, T. (2019a). Sex-biased gene expression is repeatedly masculinized in asexual females. *Nature Communications*, *10*(1), 1. <https://doi.org/10.1038/s41467-019-12659-8>
- Parker, D. J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M., & Schwander, T. (2019b). Repeated evolution of asexuality involves convergent gene expression changes. *Molecular Biology and Evolution*, *36*(2), 350–364. <https://doi.org/10.1093/molbev/msy217>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Simon, J.-C., Baumann, S., Sunnucks, P., Hebert, P. D. N., Pierre, J.-S., Le Gallic, J.-F., & Dedryver, C.-A. (1999). Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, *8*(4), 531–545. <https://doi.org/10.1046/j.1365-294x.1999.00583.x>
- Simon, J.-C., Risper, 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), 488–496. <https://doi.org/10.1016/j.crv.2010.03.003>
- Tosto, N. M., Beasley, E. R., Wong, B. B. M., Mank, J. E., & Flanagan, S. P. (2023). The roles of sexual selection and sexual conflict in shaping patterns of genome and transcriptome variation. *Nature Ecology & Evolution*, *7*(7), 7. <https://doi.org/10.1038/s41559-023-02019-7>

- Veltsos, P., Fang, Y., Cossins, A. R., Snook, R. R., & Ritchie, M. G. (2017). Mating system manipulation and the evolution of sex-biased gene expression in *Drosophila*. *Nature Communications*, *8*(1), 1. <https://doi.org/10.1038/s41467-017-02232-6>
- 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*(12), 3493–3504. <https://doi.org/10.1046/j.1365-294X.2003.01998.x>
- Whittle, C. A., & Extavour, C. G. (2019). Selection shapes turnover and magnitude of sex-biased expression in *Drosophila* gonads. *BMC Evolutionary Biology*, *19*(1), 60. <https://doi.org/10.1186/s12862-019-1377-4>
- Whittle, C. A., & Johannesson, H. (2013). Evolutionary dynamics of sex-biased genes in a hermaphrodite fungus. *Molecular Biology and Evolution*, *30*(11), 2435–2446. <https://doi.org/10.1093/molbev/mst143>
- Wilinski, D., Buter, N., Klocko, A. D., Lapointe, C. P., Selker, E. U., Gasch, A. P., & Wickens, M. (2017). Recurrent rewiring and emergence of RNA regulatory networks. *Proceedings of the National Academy of Sciences*, *114*(14), E2816–E2825. <https://doi.org/10.1073/pnas.1617777114>
- Wilson, A. C. C., Sunnucks, P., & Hales, D. F. (1997). Random loss of X chromosome at male determination in an aphid, *Sitobion near fragariae*, detected using an X-linked polymorphic microsatellite marker. *Genetics Research*, *69*(3), 233–236. <https://doi.org/10.1017/S0016672397002747>
- Wright, A. E., Fumagalli, M., Cooney, C. R., Bloch, N. I., Vieira, F. G., Buechel, S. D., Kolm, N., & Mank, J. E. (2018). Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evolution Letters*, *2*(2), 52–61. <https://doi.org/10.1002/evl3.39>

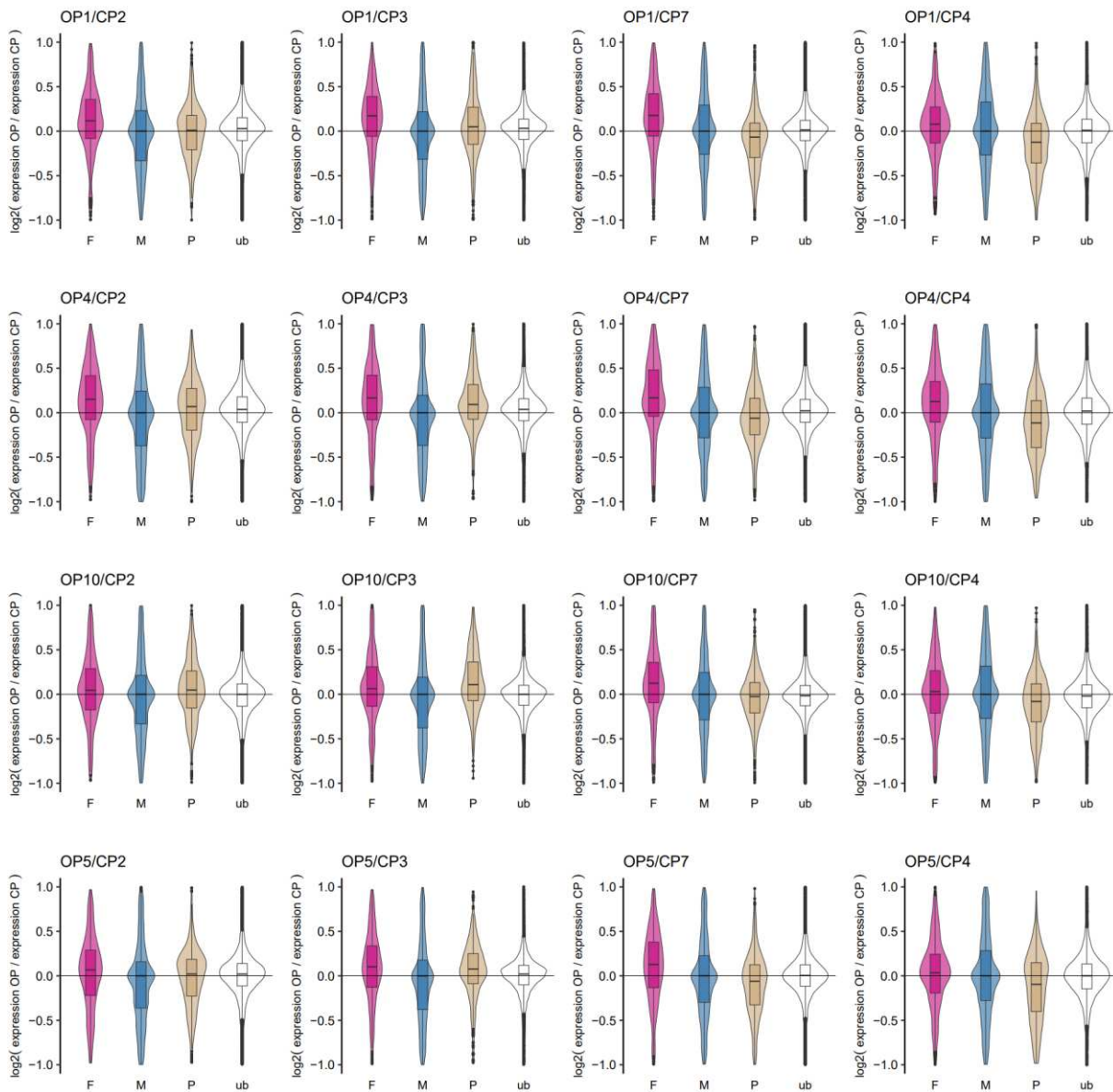
Supplementary materials for chapter 4



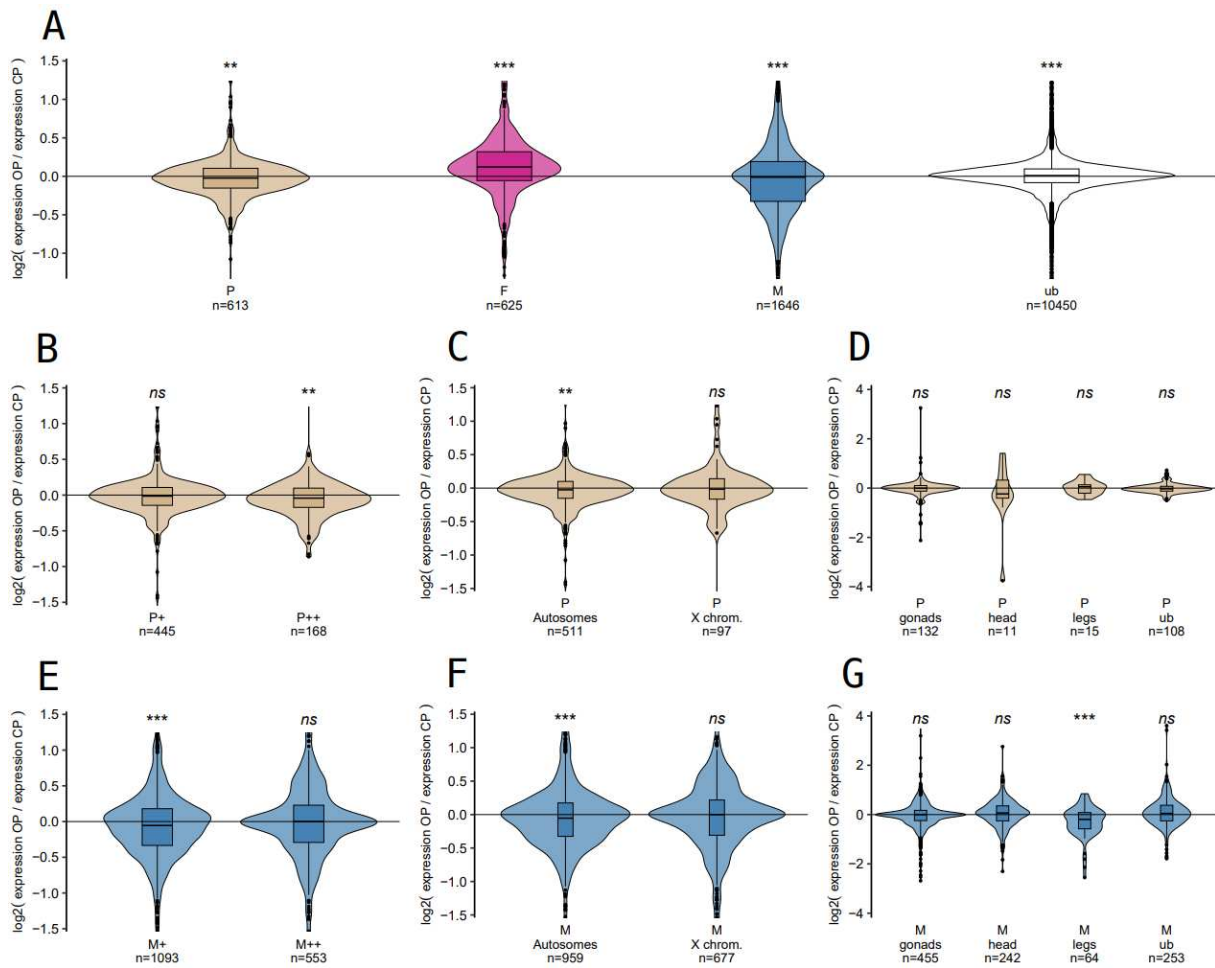
Suppl. Figure 1. Genetic relationships between the four obligate parthenogenetic (OP) and the four cyclical parthenogenetic (CP) lineages of *Acyrthosiphon pisum* used in this study. **(A)** Neighbour-joining tree built from the Nei distances between the eight lineages over 11,449 SNPs. The numbers at the nodes are bootstrap percentages (1000 replicates). **(B)** Principal Component Analysis based on 11,449 SNPs. OP lineages are shown in green and CP lineages in black.



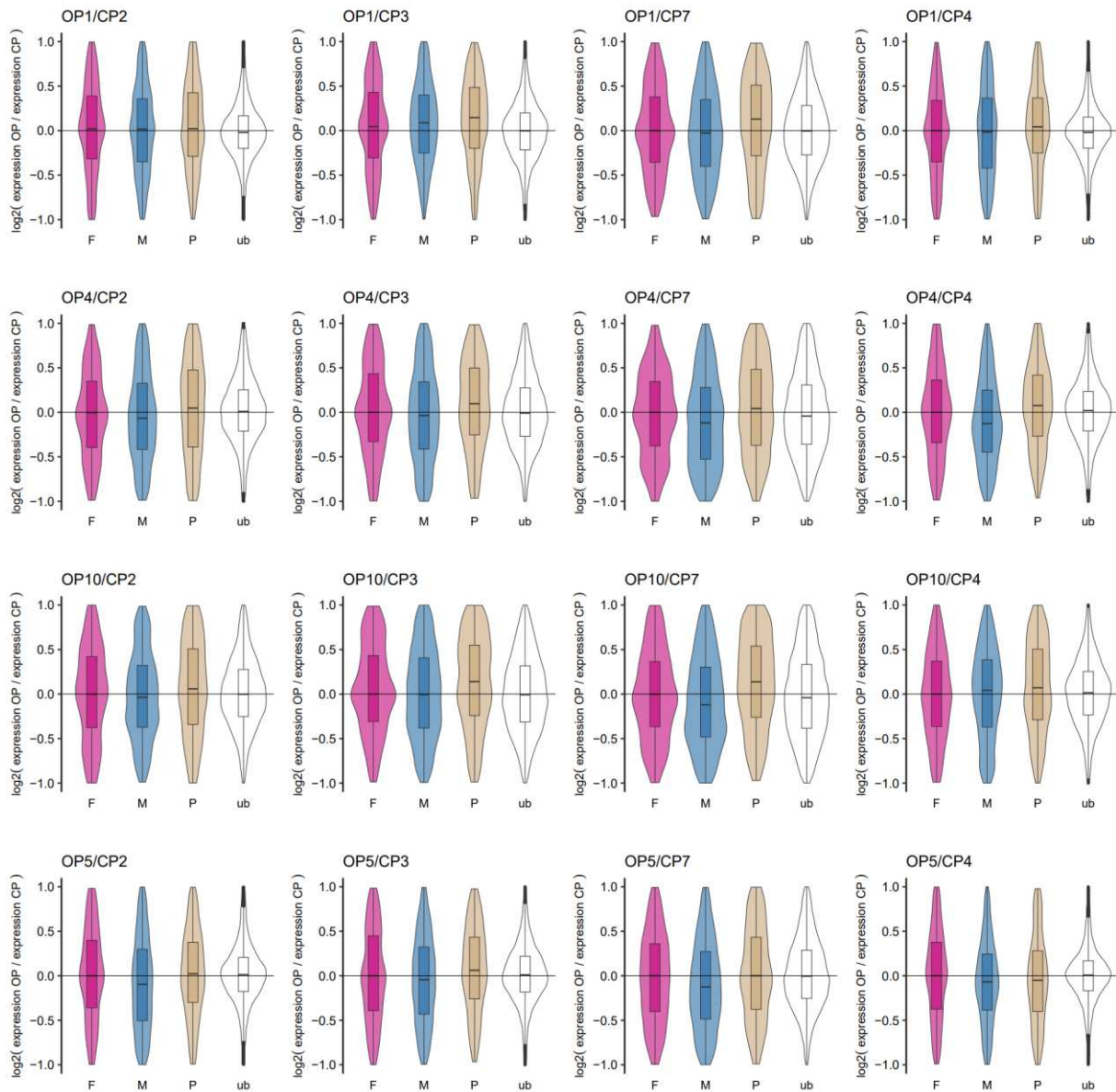
Suppl. Figure 2. Proportion of morph-biased (or unbiased) genes identified with a CP lineage of the alfalfa host race that are also morph-biased (or unbiased) according to the categorization made on the four CP lineages of the clover host race. Genes identified as morph-biased in the alfalfa host race CP lineage are predominantly biased for the same morph in the four CP lineages of the clover host race. About 81.7% of the genes identified as parthenogenetic female-biased with the outgroup also had parthenogenetic female-biased expression according to the four CP lineages from the clover host race. Of the morph-biased genes identified with the alfalfa outgroup, 74.5% of the male-biased genes and 83.7% of the sexual female-biased genes had the same morph bias of expression and 95.3% of the unbiased genes were still unbiased with the four CP lineages from the clover host race.



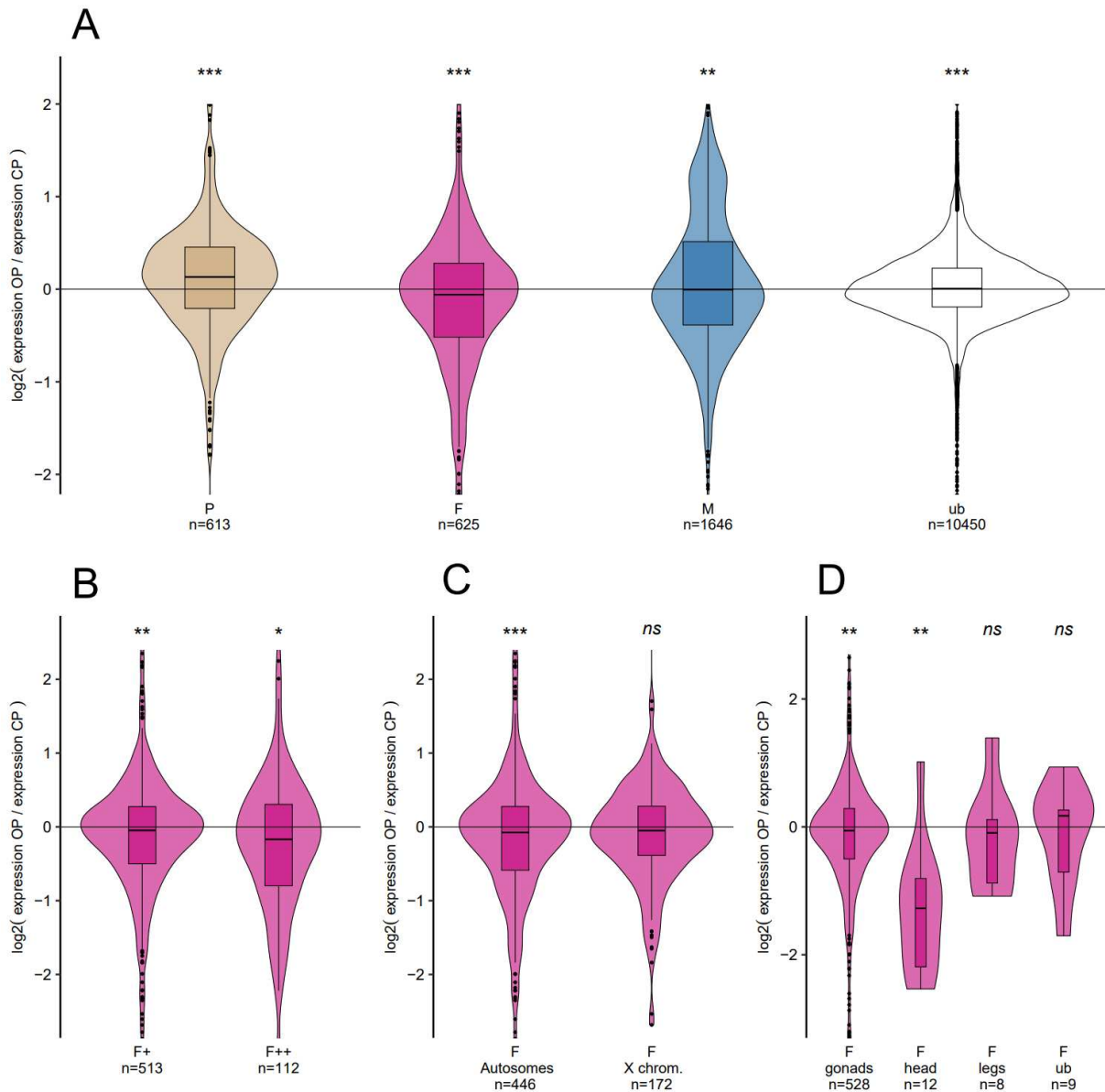
Suppl. Figure 3. Gene expression changes in parthenogenetic females between OP and CP lineages of the pea aphid, reported for each possible OP-CP lineage pairwise comparison. Genes were grouped into different classes of expression: colours discriminate between sexual female-biased genes (F, in pink), parthenogenetic female-biased genes (P, in beige), male-biased genes (M, in blue) and unbiased genes (ub, in white).



Suppl. Figure 4. Log₂-ratio of OP-to-CP gene expression in parthenogenetic females for different categories of genes. OP-to-CP expression ratio as a function of: **(A)** morph-biased status of genes, **(B, E)** degree of morph-bias (+: biased only or ++: limited) **(C, F)** chromosomal location of genes, and **(D, G)** morph- and tissue-bias status of genes. Colours discriminate between sexual female-biased genes (F, in pink), parthenogenetic female-biased genes (P, in beige), male-biased genes (M, in blue) and unbiased genes (ub, in white). The significance level of Wilcoxon signed-rank tests against 0 is indicated by asterisks (***) < 0.001, ** < 0.01, *ns* ≥ 0.05).



Suppl. Figure 5. Gene expression changes in males between OP and CP lineages of the pea aphid, reported for each possible OP-CP lineage pairwise comparison. Genes were grouped into different classes of expression: colours discriminate between sexual female-biased genes (F, in pink), parthenogenetic female-biased genes (P, in beige), male-biased genes (M, in blue) and unbiased genes (ub, in white).



Suppl. Figure 6. Log₂-ratio of OP-to-CP gene expression in males for different categories of genes. OP-to-CP expression ratio as a function of: **(A)** morph-biased status of, **(B)** degree of morph-bias (+: biased only or ++: limited) **(C)** chromosomal location of genes, and **(D)** morph- and tissue-bias status of genes. Colours discriminate between sexual female-biased genes (F, in pink), parthenogenetic female-biased genes (P, in beige), male-biased genes (M, in blue) and unbiased genes (ub, in white). The significance level of Wilcoxon signed-rank tests against 0 is indicated by asterisks (*** < 0.001, ** < 0.01, * < 0.05, ns ≥ 0.05).

SUPPLEMENTARY TABLES

Suppl. Table 1. RNAseq libraries generated for the present study. OP: obligate parthenogen; CP: cyclic parthenogen; P: parthenogenetic female; M: male; F: sexual female.

Simplified lineage ID	Lineage ID	Reproductive mode	Morph	Replicate	No. of raw paired-end reads (in M)	% mapped	Library ID on NCBI
OP1	5_12	OP	P	1	45.677583	89.3	SAMN37704917
OP1	5_12	OP	P	2	31.032626	90.5	SAMN37704918
OP1	5_12	OP	M	1	49.580856	89.3	SAMN37704915
OP1	5_12	OP	M	2	50.313871	89.5	SAMN37704916
OP4	Mn_tr_06	OP	P	1	23.84137	90.4	SAMN37704925
OP4	Mn_tr_06	OP	P	2	32.570776	91.0	SAMN37704926
OP4	Mn_tr_06	OP	M	1	52.669506	89.7	SAMN37704923
OP4	Mn_tr_06	OP	M	2	47.367056	90.0	SAMN37704924
OP5	Ta_tr_36	OP	P	1	26.630342	91.0	SAMN37704921
OP5	Ta_tr_36	OP	P	2	28.350535	90.7	SAMN37704922
OP5	Ta_tr_36	OP	M	1	58.366409	88.8	SAMN37704919
OP5	Ta_tr_36	OP	M	2	44.635683	88.4	SAMN37704920
OP10	T_tr_04	OP	P	1	23.014648	90.1	SAMN37704929
OP10	T_tr_04	OP	P	2	24.348941	90.5	SAMN37704930
OP10	T_tr_04	OP	M	1	53.950774	88.5	SAMN37704927
OP10	T_tr_04	OP	M	2	51.259362	89.6	SAMN37704928
CP2	Ar_tr_17	CP	P	1	86.018448	91.2	SAMN37704901
CP2	Ar_tr_17	CP	P	2	29.391853	91.3	SAMN37704902
CP2	Ar_tr_17	CP	M	1	44.550482	90.6	SAMN37704897
CP2	Ar_tr_17	CP	M	2	50.983626	89.7	SAMN37704898
CP2	Ar_tr_17	CP	F	1	44.909025	92.8	SAMN37704899
CP2	Ar_tr_17	CP	F	2	71.239398	93.3	SAMN37704900
CP3	Ed_tr_12	CP	P	1	29.335048	90.9	SAMN37704895
CP3	Ed_tr_12	CP	P	2	40.202976	91.7	SAMN37704896
CP3	Ed_tr_12	CP	M	1	44.748863	90.8	SAMN37704891
CP3	Ed_tr_12	CP	M	2	49.933086	90.2	SAMN37704892
CP3	Ed_tr_12	CP	F	1	54.423563	93.0	SAMN37704893
CP3	Ed_tr_12	CP	F	2	27.143991	93.0	SAMN37704894
CP4	Vm_tr_08	CP	P	1	24.090406	90.7	SAMN37704913
CP4	Vm_tr_08	CP	P	2	109.861874	90.3	SAMN37704914
CP4	Vm_tr_08	CP	M	1	44.304633	88.7	SAMN37704909
CP4	Vm_tr_08	CP	M	2	52.885104	87.6	SAMN37704910
CP4	Vm_tr_08	CP	F	1	29.713435	93.2	SAMN37704911
CP4	Vm_tr_08	CP	F	2	17.580356	92.5	SAMN37704912
CP7	Mab_tr_02	CP	P	1	26.979347	90.7	SAMN37704907
CP7	Mab_tr_02	CP	P	2	63.832718	91.0	SAMN37704908
CP7	Mab_tr_02	CP	M	1	46.386841	88.9	SAMN37704903
CP7	Mab_tr_02	CP	M	2	54.131316	86.5	SAMN37704904
CP7	Mab_tr_02	CP	F	1	46.004796	93.8	SAMN37704905
CP7	Mab_tr_02	CP	F	2	107.679125	93.4	SAMN37704906

Suppl. Table 2. Number of genes identified as being expressed preferentially in parthenogenetic females (P), sexual females (F), males (M) or unbiased based on the re-analysis of six whole-body RNAseq libraries of these three morphs (collected by Jaquiéry et al., 2013; <https://doi.org/10.1371/journal.pgen.1003690>). These genes were further subdivided into morph-biased (P+, F+ or M+) and morph-limited (P++, F++ or M++) classes, and their chromosome localisation is also given (X-chromosome, autosomes or unknown). Then, based on the re-analysis of a second RNAseq dataset of three tissues (gonads, heads and legs) in the three morphs (collected by Jaquiéry et al., 2022, <https://doi.org/10.24072/pcjournal.166>), the different types of morph-biased and morph-limited genes were further divided into sub-classes according to their expression in the different tissues. The data highlighted in beige, pink and blue correspond to the genes used in Figs. 2D and 4DG (main text) and in Suppl. Figs. 4GD and 6D.

Morph-bias status of genes - assessed with whole-body RNAseq of the 3 morphs																							
P+ and P++						F+ and F++						M+ and M++						Unbiased					
613						625						1646						10450					
P+			P++			F+			F++			M+			M++								
445			168			513			112			1093			553								
X	A	unknown chrom.	X	A	unknown chrom.	X	A	unknown chrom.	X	A	unknown chrom.	X	A	unknown chrom.	X	A	unknown chrom.	X	A	unknown chrom.			
73	368	4	24	143	1	159	347	7	13	99	0	335	751	7	342	208	3	1681	8709	60			
Tissue-bias status of genes - assessed with RNAseq from 3 tissues in the 3 morphs																							
P+			P++			F+			F++			M+			M++			unbiased					
Parthenogenetic female gonads			73			59			1			0			0			0			60		
Parthenogenetic female head			11			0			1			0			49			0			191		
Parthenogenetic female legs			2			13			1			0			3			0			4		
Sexual female gonads			0			0			433			95			2			0			68		
Sexual female head			8			0			8			4			13			0			56		
Sexual female legs			4			2			5			3			6			0			36		
Male gonads			11			4			4			6			133			322			141		
Male head			9			0			0			0			138			104			69		
Male legs			17			9			1			0			32			32			32		
Gonads (in all morphs)			42			6			39			0			10			1			1370		
Head (in all morphs)			71			2			7			0			262			5			1059		
Legs (in all morphs)			79			24			0			0			232			8			1187		
Unbiased			90			18			9			0			182			71			5164		

Suppl. Table 3. Nei genetic distances (above the diagonal) between the eight different lineages based on 11,449 SNPs. Transcriptomic Euclidean distances are shown below the diagonal, with bold beige values corresponding to those calculated between the transcriptomes of parthenogenetic females and blue values corresponding to those calculated between the transcriptomes of males.

Lineage ID	CP4	CP2	CP3	CP7	OP4	OP10	OP5	OP1
CP4	-	0.166	0.165	0.164	0.176	0.163	0.185	0.160
CP2	137.6 146.4	-	0.153	0.164	0.172	0.174	0.171	0.155
CP3	156.3 138.6	136.4 134.5	-	0.165	0.175	0.166	0.175	0.151
CP7	133.8 156.8	139.2 157.1	139.5 132.4	-	0.182	0.172	0.189	0.159
OP4	162.3 144.4	161.8 148.0	154.4 155.9	152.8 186.0	-	0.159	0.147	0.164
OP10	147.9 149.2	146.2 158.2	141.1 171.4	139.2 194.9	142.7 126.2	-	0.160	0.139
OP5	155.8 136.2	147.7 149.3	142.4 138.9	147.5 159.0	138.3 133.9	137.3 146.8	-	0.173
OP1	147.9 131.9	150.7 137.3	146.4 131.4	136.3 151.4	139.6 141.5	134.2 140.4	138.8 133.2	-

SUPPLEMENTARY TEXT

Suppl. Text 1.

This supplementary text contains the methods and results of an additional analysis aimed at investigating the differential gene expression between OP and CP lineages, using a complementary approach to that presented in the main text of the manuscript. It also includes the methods and results of an analysis investigating whether differences in gene expression could be due to differences in dosage compensation between the two types of lineages.

Materials and methods

Differential expression analyses between OP and CP lineages

Differentially expressed genes (DE) between the 8 OP and the 8 CP female libraries (in CPM) were identified using Wilcoxon rank-sum tests, as advised by (Li et al., 2022), see also Hawinkel et al. (2020). We used this test because we observed that analyses performed with statistical tools dedicated to RNA-seq analyses (notably EdgeG and DESeq2) produced erroneous results for a small fraction of the genes (which were either identified as highly significantly DE, or not DE, when obviously this was not the case). That kind of problem has already been reported (Boulain et al., 2019; Gauthier et al., 2020; Hawinkel et al., 2020; Li et al., 2022). P-values were corrected for multiple testing (FDR < 5%) using the Benjamini-Hochberg method implemented in R and only genes with a minimum absolute log₂ fold-change of 0.5 were considered as significantly DE. The same analysis was performed to identify DE genes between OP and CP males. Then, we tested whether the proportion of genes identified as DE in both males and parthenogenetic females was higher than expected by chance by performing a hypergeometric test.

Next, we tested for GO terms enrichment in the lists of DE genes in males and females against the whole set of genes. GO annotations from *A. pisum* JIC1 v1.0 genome were retrieved from AphidBase (Blast2GO on *Acyrtosiphon_pisum_JIC1_v1.0.scaffolds.braker2.gff*). GO enrichment tests were performed with topGO R package (v2.52.0, Alexa & Rahnenfuhrer, 2023) and we summarized the redundancy of GO terms with rrvgo R package (v1.12.0, Sayols, 2023) using Fisher's exact tests ($p < 0.01$). We also tested whether the two lists of genes identified as DE showed enrichment in genes previously identified as sex-biased by hypergeometric tests with the R function *phyper*.

Dosage compensation in OP and CP lineages

Here, we investigate whether possible differences in dosage compensation may also contribute to the variation in gene expression between OP and CP lineages. For that, all genes with at least one CPM in one library (out of 40) were retained. Average CPM per gene and per morph were computed for OP and CP lineages. The log₂ male-to-parthenogenetic female ratio of CPM (adding 0.05 to the numerator and the denominator to avoid division by zero) was then calculated separately for autosomal genes (n = 12,395) and X-linked genes (n = 4,554) for each type of lineage. We then tested for differences in these expression ratios between OP and CP lineages using paired Wilcoxon signed-rank tests.

Results

Differentially expressed genes between CP and OP lineages are enriched in sex-biased genes

We identified 144 DE genes between OP and CP in parthenogenetic females (79 up-regulated and 65 down-regulated in OP, Table 1 in this Suppl. Text). The set of 144 DE genes was enriched in GO terms involved in biological processes related to DNA integration and S-adenosylmethionine metabolic process (Table 2 in this Suppl. Text), though the enrichment was low. Differentially expressed genes between OP and CP parthenogenetic females were depleted in unbiased genes ($p = 1.82 \cdot 10^{-5}$, Table 1 in this Suppl. Text), but enriched in male-biased genes ($p = 0.011$) and sexual female-biased genes ($p = 0.018$). While these male-biased genes were for most of them down-regulated by OP parthenogenetic females, all but one sexual female-biased genes were up-regulated by OP parthenogenetic females. X-linked genes were more likely to be differentially expressed by OP and CP parthenogenetic females ($p = 1.39 \cdot 10^{-7}$, Table 1 in this Suppl. Text).

Table 1. Summary of enrichment tests with hypergeometric distribution on 5% FDR genes differentially expressed between OP and CP parthenogenetic females (and males, respectively), according to their morph-bias status and to their genomic location.

Expression pattern	whole genome <i>n</i> = 14,319	DE genes in parthenogenetic females (<i>n</i> = 144)			DE genes in males (<i>n</i> = 123)		
		DE genes	<i>p</i> [†]	DE genes down in OP:up in OP	DE genes	<i>p</i> [†]	DE genes down in OP:up in OP
Unbiased genes	10,450 (78.4%)	76 (66.7%)	1.82×10 ⁻⁵	37:39	69 (63.9%)	1.77×10 ⁻⁴	36:33
Sexual female-biased genes	625 (4.69%)	11 (9.65%)	0.018	1:10	10 (9.26%)	0.030	8:2
Male-biased genes	1,646 (12.3%)	23 (20.2%)	0.011	18:5	18 (16.7%)	0.11	11:7
Parthenogenetic female-biased genes	613 (4.60%)	4 (3.51%)	0.22	2:2	11 (10.2%)	0.011	2:9
	total = 13,334	total = 114[‡]			total = 108[‡]		
Chromosomal location							
X-linked genes	3,072 (21.6%)	57 (41.3%)	1.39×10 ⁻⁷	23:34	31 (25.4%)	0.18	14:17
Autosomal genes	11,123 (78.4%)	81 (58.7%)	1.39×10 ⁻⁷	39:42	91 (74.6%)	0.18	47:44
	total = 14,195	total = 138[§]			total = 122[§]		

Note: [†]Hypergeometric test *p*-value (enrichment or depletion test). [‡]The sums of the genes do not add up to the numbers given per morph (144 and 123). This is because some genes were filtered out for minimum expression in the outgroup RNA-seq dataset that was used for the identification of morph-biased genes, hence we do not have information regarding their morph bias status. [§]Six genes and one gene were not assigned to any chromosome among the DE genes for parthenogenetic females and males, respectively.

Table 2. GO terms significantly enriched at 0.01 threshold among the 144 genes identified as DE between OP and CP parthenogenetic females.

GO.ID	Term	Total annotated	Observed	Expected	P-value (Fisher exact test)
GO:0015074	DNA integration	420	9	2.49	0.00066
GO:0046500	S-adenosylmethionine metabolic process	11	2	0.07	0.00182

A total of 123 DE genes were identified between OP and CP males using the same thresholds, of which half was up-regulated and half down-regulated in OP lineages (Table 1). GO terms analyses revealed little enrichment. Main identified GO terms were DNA-templated transcription initiation and RNA 3'-end processing (Table 3). Differentially expressed genes between OP and CP males were depleted in unbiased genes ($p = 1.52 \cdot 10^{-6}$) and enriched in parthenogenetic female-biased genes ($p = 0.011$) and sexual female-biased genes ($p = 0.030$, table 1). No effect was detected for male-biased genes ($p > 0.1$). Most parthenogenetic female-biased genes were upregulated whereas most sexual female-biased genes were downregulated in OP males. A quarter of these DE genes located on the X chromosome, with no enrichment for X-linked genes among the DE genes ($p = 0.18$).

Among the 144 DE genes in parthenogenetic females and the 123 DE genes in males, 22 genes were differentially expressed in both morphs, which is higher than expected by chance ($p = 1.06 \times 10^{-21}$). The change in expression was in the same direction for all 22 genes, with 14 genes more expressed in OP than in CP lineages, and eight showing the opposite.

Table 3. GO terms significantly enriched at 0.01 threshold among the 123 genes identified as DE between OP and CP males.

GO.ID	Term	Total annotated	Observed	Expected	P-value (Fisher exact test)
GO:0006352	DNA-templated transcription initiation	92	5	0.7	0.00063
GO:0031123	RNA 3'-end processing	24	3	0.18	0.00075
GO:0031124	mRNA 3'-end processing	15	2	0.11	0.00558
GO:0071559	response to transforming growth factor beta	16	2	0.12	0.00635
GO:0071560	cellular response to transforming growth factor beta stimulus	16	2	0.12	0.00635
GO:0070126	mitochondrial translational termination	1	1	0.01	0.00759

Dosage compensation

We investigated whether possible differences in dosage compensation could account for differences in gene expression between OP and CP lineages. We found no major differences in male-to-female expression ratios between OP and CP lineages for either X or autosomes, although their distributions were statistically different due to the large sample sizes (Figure 1, this Suppl. Text).

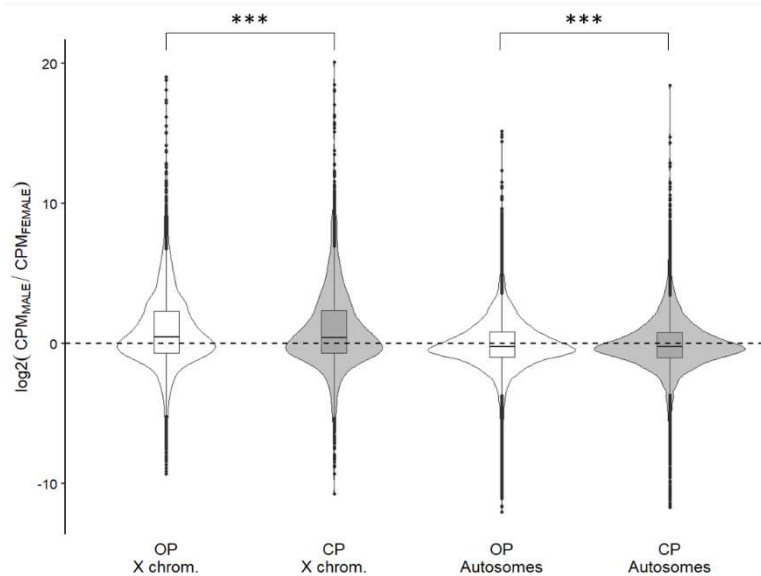


Figure 1. Log₂-ratio of male-to-female gene expression (in CPM) in OP and in CP lineages, for X-linked and autosomal genes. Expression ratios are displayed in grey for X-linked genes and in white for autosomal genes. The significance level of paired Wilcoxon signed-rank tests is indicated by asterisks (***) ($*** < 0.001$).

References

- Alexa, A., & Rahnenfuhrer, J. (2023). *TopGO: enrichment analysis for Gene Ontology* (2.52.0). <https://bioconductor.org/packages/topGO>
- Boulain, H., Legeai, F., Jaquiéry, J., Guy, E., Morlière, S., Simon, J.-C., & Sugio, A. (2019). Differential expression of candidate salivary effector genes in pea aphid biotypes with distinct host plant specificity. *Frontiers in Plant Science*, 10. <https://www.frontiersin.org/articles/10.3389/fpls.2019.01301>
- Gauthier, M., Agniel, D., Thiébaud, R., & Hejblum, B. P. (2020). dearseq: A variance component score test for RNA-seq differential analysis that effectively controls the false discovery rate. *NAR Genomics and Bioinformatics*, 2(4), lqaa093. <https://doi.org/10.1093/nargab/lqaa093>
- Hawinkel, S., Rayner, J. C. W., Bijmens, L., & Thas, O. (2020). Sequence count data are poorly fit by the negative binomial distribution. *PLOS ONE*, 15(4), e0224909. <https://doi.org/10.1371/journal.pone.0224909>
- Li, Y., Ge, X., Peng, F., Li, W., & Li, J. J. (2022). Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biology*, 23(1), 79. <https://doi.org/10.1186/s13059-022-02648-4>
- Sayols, S. (2023). rrvgo: A Bioconductor package for interpreting lists of Gene Ontology terms. *MicroPublication Biology*, 10.17912/micropub.biology.000811. <https://doi.org/10.17912/micropub.biology.000811>

Supplementary tables are available on

<https://drive.google.com/drive/folders/1ujOkzfiNIU3EWqthAqciDCXXRdrD1Y-R?usp=sharing>

Discussion

The aim of my thesis, which used aphids as a model system, was twofold. First, we wanted to understand the genetic basis of sex loss in aphids, where transitions to asexuality are common. In particular, we wanted to test whether the same loci are involved in sex loss in related or more distant aphid taxa, or whether sex loss events result from different mechanisms. We then considered the consequences of sex loss for the evolution of asexual lineages. Here we were particularly interested in testing the hypothesis that sex loss would significantly affect sex-related gene expression and phenotype, allowing us to identify a visible signature of sex loss by comparing sexual or asexual lineages of the same aphid species. I will discuss the main results of my thesis in these two parts, which together aim to contribute to the understanding of the origin and maintenance of asexual lineages, a major topic in evolutionary biology.

1. Genetic bases of sex loss

Little is known about the genomic regions involved in sex loss and even less about the genes responsible for such a change in reproductive mode. Therefore, we do not know how easy these transitions are or whether there is evolutionary convergence in the genetic basis of sex loss in related taxa. During my PhD, I investigated the genetic basis of sex loss in aphids, a group that displays reproductive polymorphism. The ancestral reproductive mode of aphids is cyclical parthenogenesis (CP, an alternation of several parthenogenetic generations and one sexual generation), but obligate parthenogenesis (OP) is frequently observed in this group.

1.1. Multiple and independent losses of sexual reproduction in aphids

In the first chapter, we have characterized a candidate region associated with sex loss in the alfalfa race of the pea aphid *Acyrtosiphon pisum*, paving the way for functional studies underway to identify the master gene(s) controlling the reproductive polymorphism in this taxon. Hugué et al. (*in revision*) analyzed the transcriptomic response to the sex-inducing environmental signal in one CP and one OP lineage of the alfalfa race of the pea aphid and identified four genes located in the candidate region that responded differently to the sex-inducing cue in the two types of lineages. *CRISPR-Cas9* genome editing is being performed by the team on these candidates to analyze their effects on reproductive phenotype when their alleles are altered or replaced. In the second chapter, we showed that the loss of sexual reproduction most likely involves different genomic regions even between closely related aphid taxa. Thus, this polymorphism in reproductive mode in the different taxa is not due to standing variation that has been maintained since their divergence, nor is the result of the spread of an *op* allele from one differentiated race to another (i.e. contagious asexuality) through hybridization. There was also no evidence for convergence neither in genomic localization nor in gene functions of the main candidate regions identified, suggesting that sex loss in aphids can be achieved in different ways. Finally, our analyses showed that there may be several loci associated with this trait in certain species.

These results suggest that the asexuality pathway in aphids is not highly constrained: different genes may be disrupted in different OP lineages, resulting in the inability to produce sexual females through various alterations. This flexibility in sex loss would maintain an array of genetically diverse

OP lineages, which may explain why asexual lineages are quite common in this group, in contrast to the majority of eukaryotes. Aphids are well suited for elucidating the genetic basis of sex loss because they show intra-specific variation in reproductive mode. However, aphid CP lineages have already acquired asexual reproduction in their life cycle, so the loss of sex in this group does not require the acquisition of a new function, but can simply occur through the loss of the ability to respond to sex-inducing cues. This could explain the easy and unconstrained loss of sex suggested here. Finally, the relatively high rates of transition to obligate parthenogenesis coupled to the existence of selective pressures that favor asexual and sexual lineages in different habitats (Frantz et al., 2006; Gilabert et al. 2009, 2014; Lehto & Haag, 2010) may allow asexuality to persist in the long term as a result of a dynamic equilibrium between the emergence of new clonal lineages and extinction due to the deleterious effects of the lack of recombination.

1.2. Asexual lineages retain evolutionary potential

The fact that aphids can repeatedly lose sexual reproduction in many different ways, both within the same species and from one species to another, raises questions about the long-term persistence of OP aphids. Asexuality may be favored in the short term and under certain environmental conditions, but it suffers from severe long-term costs, in particular impeded adaptation in the absence of recombination. Asexual aphids seem to follow the general pattern of asexual eukaryotes distribution at the tip of phylogenetic tree, with OP lineages that are most probably young (Judson & Normark, 1996; Moran, 1992). However, the fact that aphids have multiple and simple ways of losing sexual reproduction may promote the adaptability of OP lineages. It increases the genetic diversity of asexual populations and therefore the likelihood that some asexual lineages are better adapted than others to a particular habitat or to a particular selective pressure. Particularly with aphids that colonize homogeneous crops, a clone that is better adapted than the others can quickly become dominant, which may explain the prevalence of one or few OP aphid “superclones” in certain areas (Harrison & Mondor, 2011; Serteyn et al., 2021; Vorburger et al., 2003). This multitude of easy paths to asexuality hence leaves more room for selection to act on asexual genotypes. In addition, the short generation time and the large size of asexual populations make it more likely that potentially beneficial mutations will occur in their genomes, and co-adapted gene complexes that are not disrupted by recombination would efficiently increase in frequency. Accordingly, rates of phenotypic evolution can be surprisingly high in parthenogenetic organisms (Lynch & Gabriel, 1983). The absence of meiosis in asexual aphid lineages also relaxes constraints on chromosomal arrangement and may accelerate genome evolution, with chromosomal rearrangements being particularly frequent due to their holocentric chromosomes (Normark, 1999). Some chromosomal rearrangements can be selectively advantageous and quickly inherited through parthenogenesis in asexual lineages. For instance, an insecticide resistance was associated with a chromosomal translocation in asexual lineages of *Myzus persicae* (Denholm et al., 1998). Contagious asexuality can also affect the evolutionary potential of OP lineages. Indeed, the mating between an OP male and a sexual female from a CP lineage will produce CP offspring, heterozygous at the reproductive mode locus given that the *op* allele is recessive (at least in the *A. pisum* alfalfa race, Jaquiéry et al., 2014). *Op* recessive allele will therefore spread across CP lineages. Then, mating between males carrying the *op* allele and CP sexual females heterozygous at the locus (*op/cp*) can create new OP lineages, with a genome purged of deleterious mutations. Contagious asexuality is strongly supported by the low global genetic differentiation (measured with F_{ST}) between

OP and CP populations in alfalfa and pea races of the pea aphid, and was also observed in other aphid species (Halkett et al., 2005) and in *Daphnia* (Paland et al., 2005). Contagious asexuality enables asexual alleles to escape mutational load of asexual lineages. This thus favors the frequent emergence of well fit OP lineages, which could eventually replace the older less fit OP lineages.

1.3. Perspectives

In contrast to our previous work on alfalfa race (Jaquiéry et al., 2014, and chapter 1), we could not clearly identify candidate regions in *A. pisum* clover and pea races, nor in *Myzus persicae*. However, this conclusion needs further support, as we indicated in Chapter 2 that our study requires complementary analysis and points to some limitations that need to be addressed in the near future. Further studies based on QTL approaches for the detection of candidate regions for the control of sex loss would be of great interest. It would be interesting to further explore the possible role of symbionts in the reproduction of aphids and test for a difference in symbiont composition between the OP and CP populations. In contrast to other arthropod species (e.g. Werren et al., 2008), symbionts have never been shown to suppress the sexual reproduction in aphids (Simon et al., 2007, 2011). Nevertheless, some may induce male-killing (Simon et al., 2011) or influence the host's resistance to cold, which can confer an advantage to OP lineages that are cold-sensitive because of their inability to produce overwintering eggs. Our preliminary results suggest that OP and CP lineages in the pea race of *A. pisum* may carry different strains of *Serratia*, a facultative symbiont known to affect aphid thermal resistance (Pons et al., 2022; Tougeron & Iltis, 2022). This hypothesis could be confirmed by F_{ST} scans on the genome of the symbionts (rather than on the aphid nuclear genome, as done in chapter 2). If these analyses confirm an association, it would then be possible to experimentally manipulate the symbiotic composition of the aphid, for example by introducing a *Serratia* lineage originally associated with an OP lineage into a CP lineage, and observe whether this induces a better resistance to cold or a change in its reproductive mode. Finally, surveys on the phylogenetic relationships between the candidate X-linked region in the alfalfa race and this region in the different race of the pea aphid complex could validate a possible hybrid origin of the *op* allele in alfalfa race from the pea race. A more straightforward validation would be to cross a CP lineage of the alfalfa race with a CP lineage of the pea race to check whether the F2 hybrids homozygous for the pea allele in the candidate region are OP and the others are CP.

2. Evolutionary consequences of sex loss

The second main objective of my thesis was to assess the consequences of sex loss on gene expression and phenotypic evolution. Three main effects are expected following sex loss. First, in asexual apomictic populations, lack of meiotic recombination leads to high genetic linkage, which would ultimately reduce the overall effectiveness of selection along with an inevitable accumulation of deleterious mutations in asexual species (“Muller’s ratchet”; Muller, 1964). One of the expected effects of absence of recombination that this thesis intended to test was an increase in the variance of gene expression or phenotype in non-recombinant lineages. Second, we looked for evidence of relaxed selection on sexual morphs and especially males, as expected after the shift from a sexually

reproducing species with males and females to an asexual all-female species. Third, the absence of sexual morphs that results from transition to asexuality reduces genetic conflict that existed between males and females due to the constraint of a shared genome. The effect of reducing intra-locus conflict in asexual lineages has been studied in species that have shifted from obligate sexual to obligate asexual reproduction, but the effect of loss of sexual conflict is confounded with the evolution of a new morph in such taxa (Huylmans et al., 2021; Parker et al., 2019). Here we aimed to study the evolutionary consequences of sex loss on the transcriptome of aphids that switched from cyclic to obligate parthenogenesis, which are free of this bias.

2.1. Transcriptomic and phenotypic changes in asexual lineages are associated with sex loss

Due to geographic isolation with CP, the few males produced by OP lineages should have little opportunity to mate, thus are supposedly dead ends regarding gene transmission. Male-related traits in OP lineages are hence expected to decay. In chapter 3, we used a large collection of different *A. pisum* lineages with contrasting reproductive modes to investigate the evolution of male-related traits following sex loss. We showed that male production in OP lineages was greatly reduced as expected, while male fertility in OP lineages was surprisingly only slightly reduced. In chapter 4, we analyzed gene expression in parthenogenetic females and males from several OP and CP pea aphid lineages to test the prediction that the disappearance of sexual conflict in OP lineages would result in shifts of gene expression towards parthenogenetic female optimum. As expected, OP males more highly express parthenogenetic female-biased genes relative to CP males. However, they surprisingly express testis-specific genes at higher rate too. Contrary to predictions, OP parthenogenetic females do not overexpress parthenogenetic female-biased genes, but they do systematically overexpress sexual female-biased genes and especially those normally expressed in sexual female ovaries, while we expected them to be less expressed. As in the two others studies that have investigated the evolutionary consequences of sex loss at the transcriptomic level (Huylmans et al., 2021; Parker et al., 2019), the empirical observations do not fully match the theoretical predictions.

2.2. Deciphering the nature and timeline of evolutionary forces underlying these changes

Knowing which evolutionary forces are responsible for the observed patterns, or how quickly these forces are acting, is challenging. Nevertheless, we can try to estimate the relative contribution of the different selective forces on our results, namely positive selection or relaxed selection with slow drift-driven effects, but also consider the influence of the absence of recombination.

First, contrary to one of our working hypotheses, some of the phenotypic and transcriptomic changes we observed can be explained by the action of positive selection. In OP lineages, males are generally still functional as confirmed in chapter 3 and showed in earlier works on different aphid species (Blackman, 1974; Dedryver et al., 2013, 2019; Jaquiéry et al., 2014). Reproductive success of OP males may also be under positive selection, with OP lineages that produce functional males being favored by selection. Maintaining male functionality would result from males having some opportunities to pass on their genes to the next generation by mating with CP sexual females. Mating

between OP males and CP females may occur in the wild, especially in regions with intermediate climates that typically favor either type of lineages depending, among other things, on the rigor of the winter (e.g. Halkett et al., 2008; Rispe et al., 1998). In our study (chapters 3 and 4), the OP and CP lineages are less likely to meet because they were sampled in areas at least 480 km apart, but the intermediate regions may contain lineages with contrasting reproductive modes with which mating may be possible. As highlighted in chapter 2, the OP lineages from the *A. pisum* clover race we used in our studies produced more winged males than CP lineages. Although this male phenotype may simply be randomly associated with the OP phenotype, this could be a footprint of selection acting on males and favoring winged males, which will be able to disperse to meet sexual females. If this is the case, it would explain the reproductive success of OP males, which are able to fertilize CP females even in competition with CP males, despite a slight decrease that could be explained by deleterious mutations accumulating throughout the non-recombining OP genome. Clarifying whether male winged phenotype in OP-selecting environments is an adaptation that favors the transmission of OP alleles through mating with distant CP lineages would help estimate the strength of selection on OP males and hence their mating opportunities. In order to do so, we could not only test whether it is functional, i.e. whether OP males can actually fly, but also whether winged males produced by OP lineages are just as frequent in intermediate regions where OP and CP lineages coexist, or whether, as expected if this trait is adaptive, they are just as rare as winged males of CP lineages. In addition to the ability of OP males to meet and mate with sexual females, another important factor affecting the efficiency of OP alleles transmission through contagious asexuality is the ability of eggs fathered by OP males to hatch successfully and the ability of OP offspring to compete with CP offspring. Competitiveness at birth was suggested for clones derived from crosses between OP and CP lineages of different aphid species hatched in average about a week earlier than their CP sibs, which gives them a demographic advantage at the start of the growth season (Dedryver et al., 2019). Besides, if the changes in gene expression that we observe are the result of positive selection (which could notably be the case for the very convergent changes in parthenogenetic females), then this would be another indication of the evolving capacities of asexual lineages. Lineages with previously sexual gonad-specific genes that acquired new functions or lineages with parthenogenetic females expressing sexual gonad-specific genes in a more optimal manner would be favored by selection. It would hence give us an idea of the extent to which gene expression in sexual lineages is actually constrained by intra-locus conflict.

Second, some results alternatively support the hypothesis of relaxed selection in OP lineages of aphids. As we expected, OP males more highly express parthenogenetic female-biased genes relative to CP males, which is the first evidence for relaxed selection on sex-biased genes in males of asexual lineages. Besides, the slight reduction of male fertility in OP lineages can also result, instead of positive selection, of slow drift-driven decay. If male production, that is clearly reduced in these lineages, is being counter-selected, this indicates that males are costly and maladaptive and suggests that their opportunity to transmit their genes is low. Apart from pleiotropic effects, male fertility maintenance would be in that case best explained by relaxed selection. Male production being quickly reduced, OP lineages that rarely produce non-functional males would not have a larger selective advantage than OP lineages that rarely produce functional males. Consequences of relaxed selection on a trait are not easily predictable. Indeed, they depend on a variety of factors that can influence the evolution of the trait value, in particular direct fitness effects of the expression of the trait (e.g. costs and benefits of OP males being functional), but also indirect effects through trait interaction (e.g. costs and benefits of OP males being produced).

Third, we wanted to assess the extent to which the lack of recombination affects the evolution of gene expression. Yet, the study of the consequences of the lack of genome recombination in asexual lineages, which should lead to independent lineage evolution and greater expression divergence between OP than between CP lineages, has not yielded convincing results, as no effect was detected in chapter 4. Including more lineages, or older lineages, might yield different results and allow for better estimation of transcriptomic divergence.

One question that arises is how fast these forces act, and hence how fast evolutionary changes occur. To better understand the timescale of the evolution of asexual lineages, knowledge of the time elapsed since sex loss would be particularly valuable. Yet, little is known about the age of asexual lineages in aphids. Evidence for short lifespan of asexual genotypes has been reported in *Daphnia pulex* (Tucker et al., 2013), with their genomes showing signs of limited longevity i.e. less than 20 years. The supposedly high turnover of asexual lineages and low longevity of asexual aphid genotypes (Delmotte et al., 2002; Vorburger, 2006) would result in young contemporary OP lineages that could replace less fit ones. The effect of the age of sex loss may be partially offset by mechanisms that reduce the rate or cost of mutation accumulation, such as efficient DNA repair or, as proposed in aphids (Normark, 1999), slow molecular evolution. Experimental estimation of the mutation rate in the parthenogenetic part of the life cycle in the pea aphid indeed found a low value (2.7×10^{-10} , Fazalova et al., 2020). Even without knowing the exact age of the asexual species being studied, it would already be of great interest to work with a panel of species or lineages that are more or less young relatively to each other, in order to understand whether age influences the observed pattern and ultimately decipher the timeline of evolutionary forces on asexual lineages. The relative age of OP lineages can be assessed by the degree of divergence from the most closely related CP lineages, which are most likely to be related to their sexual ancestor, but would require extensive sampling of lineages to improve the accuracy of estimates. Given the multiple and spontaneous sex loss events in aphids, it is likely that at least some OP aphid lineages differ in age. Extensive sampling, particularly in intermediate regions where OP and CP coexist, may help to identify a gradient from young to old asexual lineages, which would help to model the evolution of their traits supposedly under relaxed selection.

2.3. Perspectives

To better understand the contribution of the different evolutionary forces acting on lineages that have recently lost sex and which could explain our results, several different approaches could be developed. First, under the hypothesis that the gene expression changes we observe are adaptive, we expect to see the same pattern in other more distantly related species that independently lost sex. A comparative transcriptomic analysis with other species where cyclically parthenogenetic and obligately parthenogenetic populations coexist would be a promising avenue. From this point, we could compare gene expression in parthenogenetic females between different OP and CP species to see if asexuality is associated with convergent shifts in gene expression first and, in particular, whether the absence of intra-locus conflict would systematically lead to the pattern we observed in *A. pisum*, i.e. OP parthenogenetic females more highly expressing sexual female-biased genes. This would be evidence that the observed changes are indeed due to the absence of intra-locus conflict and the relaxation of selection on sex-biased genes, leading to that particular expression pattern which may therefore be optimal for and selected for in parthenogenetic females. Aphids are convenient for this comparative analysis because this group comprises several pairs of aphid species that are closely

related (belonging to the same genus) but differ in their reproductive mode. During my PhD, we initiated this analysis by generating the necessary transcriptomic data, but there was not enough time to check whether the shifts in expression observed at the intra-specific level in *A. pisum* clover race were also found at the inter-specific level.

An alternative and exciting perspective to gain further insight into the evolutionary forces that drive the evolution of sex-biased genes after sex loss would be to focus on their DNA sequences. Although it was in the original plan of the thesis, we had not enough time to develop this part. More specifically, we could examine the patterns of intraspecific polymorphism and interspecific divergence in sex-biased genes relative to parthenogenetic female-biased genes and compare them between OP and CP aphid lineages. First, we could characterize the genomic polymorphism (pN, pS and pN/pS) at the intra-specific scales for *A. pisum* and *M. persicae*, the two species for which we already have population genetic data (see chapters 1 and 2). We expect 1) an overall higher pN/pS ratio for all categories of genes carried by OP genomes, due to the inevitable accumulation of deleterious mutations in asexual genomes, 2) a higher pN/pS ratio for sex-biased genes in both OP and CP genomes, as genes expressed in the rare (or absent) morphs are predicted to accumulate deleterious mutations at a higher rate than genes always expressed (unbiased genes) or expressed in the most common morph (parthenogenetic female-biased genes) and 3) an interaction effect, with higher pN/pS for male-biased (and sexual female-biased) genes in OP than in CP populations, and a smaller difference for the genes biased in parthenogenetic females. On another evolutionary scale, we can also work on different aphid species that are exclusively CP or exclusively OP and in which loss of sex is thought to have arisen independently, which has the potential to show more contrasted effects. Although few exclusively OP aphid species have been described, *Aphis gossypii* (Carletto et al., 2009), *Acyrtosiphon kondoi* (MacKay & Lamb, 1988), *Myzus ascalonicus* (Blackman & Eastop, 2000) and *Rhopalosiphum maidis* (Blackman & Eastop, 2000), for example, are known to be essentially OP worldwide and could be used for this purpose.

Here too, if sex-biased genes evolve under relaxed selection in OP species, we expect an accumulation of deleterious mutations that translates into higher rate of substitution (dN/dS) for orthologous sex-biased genes shared between sexual and asexual species.

Another prediction of the lack of the recombination on gene expression is a haploidization of expression due to an enhancer divergence process in OP lineages, a hypothesis that I did not test because of time constraints. Fyon & Lenormand (2018) predict that haploidization should evolve whenever homolog chromosomes are sufficiently isolated and heterozygosity is high, i.e., low rates of recombination or gene conversion. We could hence assess allele-specific expression on RNA-seq data in OP and CP lineages, and test whether allele-specific expression is stronger or more frequent in OP than in CP lineages.

3. Conclusion

The results of my thesis show that sexual reproduction can be lost in different ways in aphids, even at the intra-specific level, and that transitions to asexuality lead to different changes in selective regimes with a wide range of effects on the evolution of asexual lineages, including consequences visible at the phenotypic and transcriptomic level. This thesis highlights the importance of reproductive mode in shaping the evolution of organisms, their genomes and traits.

4. References

- Blackman, R. L. (1974). Life-cycle variation of *Myzus persicae* (Sulz.) (*Hom., Aphididae*) in different parts of the world, in relation to genotype and environment. *Bulletin of Entomological Research*, 63(4), 595–607. <https://doi.org/10.1017/S0007485300047830>
- Blackman, R. L., & Eastop, V. F. (2000). *Aphids on the world's crops: An identification and information guide*. John Wiley & Sons Ltd. <https://www.cabdirect.org/cabdirect/abstract/20001106823>
- Carletto, J., Lombaert, E., Chavigny, P., Brévault, T., Lapchin, L., & Vanlerberghe-Masutti, F. (2009). Ecological specialization of the aphid *Aphis gossypii* Glover on cultivated host plants. *Molecular Ecology*, 18(10), 2198–2212. <https://doi.org/10.1111/j.1365-294X.2009.04190.x>
- Dedryver, C.-A., Bonhomme, J., Le Gallic, J.-F., & Simon, J.-C. (2019). Differences in egg hatching time between cyclical and obligate parthenogenetic lineages of aphids. *Insect Science*, 26(1), 135–141. <https://doi.org/10.1111/1744-7917.12493>
- Dedryver, C.-A., Le Gallic, J.-F., Mahéo, F., Simon, J.-C., & Dedryver, F. (2013). The genetics of obligate parthenogenesis in an aphid species and its consequences for the maintenance of alternative reproductive modes. *Heredity*, 110(1), 1. <https://doi.org/10.1038/hdy.2012.57>
- Delmotte, F., Leterme, N., Gauthier, J.-P., Rispe, C., & Simon, J.-C. (2002). Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology*, 11(4), 711–723. <https://doi.org/10.1046/j.1365-294X.2002.01478.x>
- Denholm, I., Pickett, J. A., Devonshire, A. L., Devonshire, A. L., Field, L. M., Foster, S. P., Moores, G. D., Williamson, M. S., & Blackman, R. L. (1998). The evolution of insecticide resistance in the peach–potato aphid, *Myzus persicae*. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1376), 1677–1684. <https://doi.org/10.1098/rstb.1998.0318>
- Frantz, A., Plantegenest, M., & Simon, J.-C. (2006). Temporal habitat variability and the maintenance of sex in host populations of the pea aphid. *Proceedings of the Royal Society B: Biological Sciences*, 273(1603), 2887–2891. <https://doi.org/10.1098/rspb.2006.3665>
- Fyon, F., & Lenormand, T. (2018). Cis-regulator runaway and divergence in asexuals. *Evolution*, 72(3), 426–439. <https://doi.org/10.1111/evo.13424>
- Gilbert, A., Simon, J.-C., Dedryver, C.-A., & Plantegenest, M. (2014). Do ecological niches differ between sexual and asexual lineages of an aphid species? *Evolutionary Ecology*, 28(6), 1095–1104. <https://doi.org/10.1007/s10682-014-9730-y>
- Gilbert, A., Simon, J.-C., Mieuze, L., Halkett, F., Stoeckel, S., Plantegenest, M., & Dedryver, C.-A. (2009). Climate and agricultural context shape reproductive mode variation in an aphid crop pest. *Molecular Ecology*, 18(14), 3050–3061. <https://doi.org/10.1111/j.1365-294X.2009.04250.x>
- Halkett, F., Plantegenest, M., Bonhomme, J., & Simon, J.-C. (2008). Gene flow between sexual and facultatively asexual lineages of an aphid species and the maintenance of reproductive mode variation. *Molecular Ecology*, 17(12), 2998–3007. <https://doi.org/10.1111/j.1365-294X.2008.03798.x>
- Halkett, F., Plantegenest, M., Prunier-Leterme, N., Mieuze, L., Delmotte, F., & Simon, J. C. (2005). Admixed sexual and facultatively asexual aphid lineages at mating sites. *Molecular Ecology*, 14(1), 325–336. <https://doi.org/10.1111/j.1365-294X.2004.02358.x>
- Harrison, J. S., & Mondor, E. B. (2011). Evidence for an invasive aphid “superclone”: extremely low genetic diversity in oleander aphid (*Aphis nerii*) populations in the Southern United States. *PLOS ONE*, 6(3), e17524. <https://doi.org/10.1371/journal.pone.0017524>
- Huguet, M. D., Robin, S., Hudaverdian, S., Tanguy, S., Leterme-Prunier, N., Cloteau, R., Baulande, S., Legoix-Né, P., Legeai, F., Simon, J.-C., Jaquiéry, J., Tagu, D., & Le Trionnaire, G. Transcriptomic analyses of the genetic bases of sex loss in the pea aphid [*Manuscript submitted for publication*].

- Huylmans, A. K., Macon, A., Hontoria, F., & Vicoso, B. (2021). Transitions to asexuality and evolution of gene expression in *Artemia* brine shrimp. *Proceedings of the Royal Society B: Biological Sciences*, 288(1959), 20211720. <https://doi.org/10.1098/rspb.2021.1720>
- Jaquiéry, J., Stoeckel, S., Larose, C., Nouhaud, P., Rispe, C., Mieuzet, L., Bonhomme, J., Mahéo, F., Legeai, F., Gauthier, J.-P., Prunier-Leterme, N., Tagu, D., & Simon, J.-C. (2014). Genetic control of contagious asexuality in the pea aphid. *PLOS Genetics*, 10(12), 12. <https://doi.org/10.1371/journal.pgen.1004838>
- Judson, O. P., & Normark, B. B. (1996). Ancient asexual scandals. *Trends in Ecology & Evolution*, 11(2), 41–46. [https://doi.org/10.1016/0169-5347\(96\)81040-8](https://doi.org/10.1016/0169-5347(96)81040-8)
- Lehto, M. P., & Haag, C. R. (2010). Ecological differentiation between coexisting sexual and asexual strains of *Daphnia pulex*. *Journal of Animal Ecology*, 79(6), 1241–1250. <https://doi.org/10.1111/j.1365-2656.2010.01726.x>
- Lynch, M., & Gabriel, W. (1983). Phenotypic evolution and parthenogenesis. *The American Naturalist*, 122(6), 745–764. <https://doi.org/10.1086/284169>
- Mackay, P. A., & Lamb, R. J. (1988). Genetic variation in asexual populations of two aphids in the genus *Acyrtosiphon*, from an Australian lucerne field. *Entomologia Experimentalis et Applicata*, 48(2), 117–125. <https://doi.org/10.1111/j.1570-7458.1988.tb01154.x>
- Moran, N. A. (1992). *The evolution of aphid life cycles*.
- Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1), 2–9. [https://doi.org/10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8)
- Normark, B. B. (1999). Evolution in a putatively ancient asexual aphid lineage: recombination and rapid karyotype change. *Evolution*, 53(5), 1458–1469. <https://doi.org/10.1111/j.1558-5646.1999.tb05410.x>
- Paland, S., Colbourne, J. K., & Lynch, M. (2005). Evolutionary history of contagious asexuality in *Daphnia pulex*. *Evolution*, 59(4), 800–813. <https://doi.org/10.1111/j.0014-3820.2005.tb01754.x>
- Parker, D. J., Bast, J., Jalvingh, K., Dumas, Z., Robinson-Rechavi, M., & Schwander, T. (2019). Sex-biased gene expression is repeatedly masculinized in asexual females. *Nature Communications*, 10(1), 1. <https://doi.org/10.1038/s41467-019-12659-8>
- Pons, I., Scieur, N., Dhondt, L., Renard, M.-E., Renoz, F., & Hance, T. (2022). Pervasiveness of the symbiont *Serratia symbiotica* in the aphid natural environment: Distribution, diversity and evolution at a multitrophic level. *FEMS Microbiology Ecology*, 98(1), fiac012. <https://doi.org/10.1093/femsec/fiac012>
- Rispe, C., Pierre, J.-S., Simon, J.-C., & Gouyon, P.-H. (1998). Models of sexual and asexual coexistence in aphids based on constraints. *Journal of Evolutionary Biology*, 11(6), 685–701. <https://doi.org/10.1046/j.1420-9101.1998.11060685.x>
- Serteyn, L., Bosquee, E., Delatour, A., Rubio-Melendez, M. E., Vega-Muñoz, K. V., Ramirez, C. C., & Francis, F. (2021). Multi-approach comparative study of the two most prevalent genotypes of pea aphid *Acyrtosiphon pisum* (Hemiptera: Aphididae) in Chile. *Entomological Science*, 24(1), 55–67. <https://doi.org/10.1111/ens.12436>
- Simon, J.-C., Boutin, S., Tsuchida, T., Koga, R., Gallic, J.-F. L., Frantz, A., Outreman, Y., & Fukatsu, T. (2011). Facultative symbiont infections affect aphid reproduction. *PLOS ONE*, 6(7), e21831. <https://doi.org/10.1371/journal.pone.0021831>
- Simon, J.-C., Sakurai, M., Bonhomme, J., Tsuchida, T., Koga, R., & Fukatsu, T. (2007). Elimination of a specialised facultative symbiont does not affect the reproductive mode of its aphid host. *Ecological Entomology*, 32(3), 296–301. <https://doi.org/10.1111/j.1365-2311.2007.00868.x>
- Tougeron, K., & Iltis, C. (2022). Impact of heat stress on the fitness outcomes of symbiotic infection in aphids: A meta-analysis. *Proceedings of the Royal Society B: Biological Sciences*, 289(1971), 20212660. <https://doi.org/10.1098/rspb.2021.2660>

- Tucker, A. E., Ackerman, M. S., Eads, B. D., Xu, S., & Lynch, M. (2013). Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. *Proceedings of the National Academy of Sciences*, *110*(39), 15740–15745. <https://doi.org/10.1073/pnas.1313388110>
- Vorburger, C. (2006). Temporal dynamics of genotypic diversity reveal strong clonal selection in the aphid *Myzus persicae*. *Journal of Evolutionary Biology*, *19*(1), 97–107. <https://doi.org/10.1111/j.1420-9101.2005.00985.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*(12), 3493–3504. <https://doi.org/10.1046/j.1365-294X.2003.01998.x>
- Werren, J. H., Baldo, L., & Clark, M. E. (2008). *Wolbachia*: Master manipulators of invertebrate biology. *Nature Reviews Microbiology*, *6*(10), 10. <https://doi.org/10.1038/nrmicro1969>

Annex

Annex 1. Poster presented at the Jacques Monod conference in Roscoff (“Sex unfolded: sex, asex, sexes”) in September 2023.

Evolution of gene expression following sex loss in the pea aphid

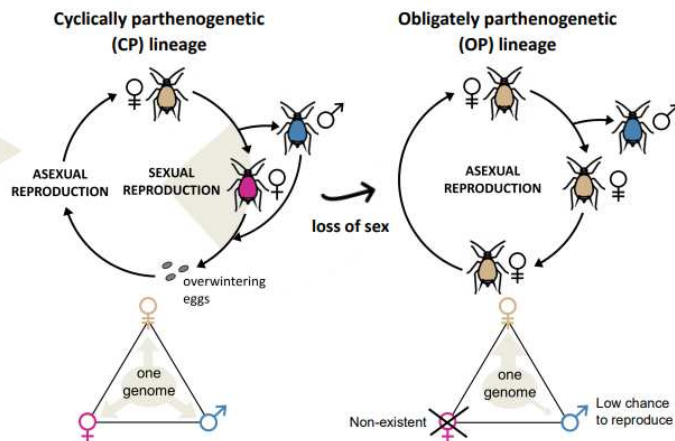
Hélène Defendini¹, Nathalie Prunier-Leterme¹, Stéphanie Robin², Jean-Christophe Simon¹, Julie Jaquiéry¹

¹ UMR 1349 IGEPP, INRAE, Institut Agro, Université de Rennes, 35653 Le Rheu, France
² UMR 1349 IGEPP, INRAE, Institut Agro, Université de Rennes, 35000 Rennes, France

Sexual conflict is widespread as males and females share one genome but exhibit different phenotypes. Sexually dimorphic gene expression may help reduce conflict, but many genes may still be suboptimally expressed due to unresolved tensions between sexes. Asexual lineages lack such conflict, making them relevant models for understanding how sexual conflict influences gene expression. We investigate the evolution of sexual conflict following sex loss by contrasting the gene expression patterns of sexual and asexual lineages in the pea aphid.

Acyrtosiphon pisum shows an intraspecific variation in reproductive mode associated with climate, with cold-resistant CP lineages displaying the ancestral reproductive mode and derived cold-sensitive OP lineages producing no sexual females but still males.

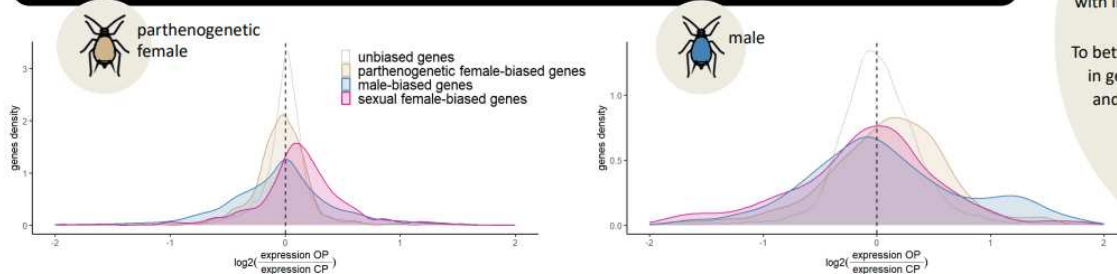
As OP and CP lineages are geographically separated, OP males have limited reproductive opportunity.



Sexual conflict may exist between the 3 morphs required to complete the annual cycle of CP lineages (Jaquiéry *et al.* 2013 Plos Genetics 9:e1003690) but this conflict should be reduced in OP lineages.

Hypothesis

Relaxed sexual conflict in OP lineages would result in shifts in gene expression toward the parthenogenetic female optimum. We predicted that parthenogenetic female-biased genes would be more highly expressed in OP morphs (parthenogenetic females and males) than CP morphs.



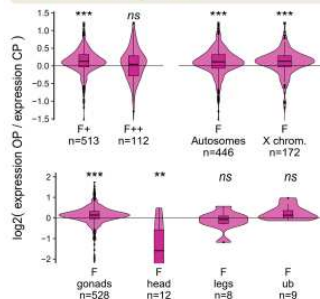
Materials & methods

We performed whole body RNA-seq for the morphs in 4 OP and 4 CP *A. pisum* lineages.

We identified morph-biased genes (>70% of reads in 1 morph) with independent RNA-seq data. To better understand the changes in gene expression between OP and CP morphs, we subdivided morph-biased genes by chromosomal location, by the degree of expression bias (+ or ++) and by the tissue in which they were most expressed.

Unexpectedly, OP parthenogenetic females overexpressed sexual female-biased genes.

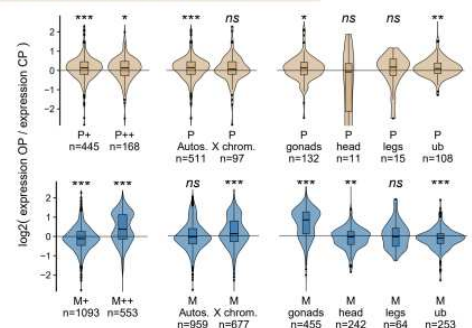
As expected, OP males overexpressed parthenogenetic female-biased genes and more surprisingly male-biased genes.



OP parthenogenetic females overexpressed genes normally expressed at high rate in the ovaries of sexual females.

Parthenogenetic female-biased genes more expressed by OP males were mostly on autosomes.

OP males highly expressed male-biased genes that were more specifically expressed in the testes, which additionally tended to be male-limited (++) and more frequent on the X chromosome (Jaquiéry *et al.* 2022, Peer Community J. 10.24072/pcjournal.166).



Significance evaluated with Wilcoxon signed-rank test

Conclusion

We observed that the expression of genes in males from asexual lineages tended towards the parthenogenetic female optimum, as predicted by theory. Surprisingly, males and parthenogenetic females of asexual lineages overexpressed genes normally found in the ovaries and testes of sexual morphs. These changes in gene expression in asexual lineages may arise from the relaxation of selection or the dysregulation of gene networks otherwise used in sexual lineages.



Acknowledgments: This work was supported by grants from the French Research Agency (ANR-09-GENM-017-001 SexAphid and ANR-11-BSV7-005-01 Mecadap), the INRAE-SPE Department (AAP GenAsex, AAP CycleEvol and half a PhD grant for HD), the Region Bretagne (half a PhD grant for HD).

Title: Genetic bases and evolutionary consequences of sex loss in aphids

Keywords: Parthenogenesis; transition to asexuality; clonal evolution; trait decay; gene expression; sexual conflict

Abstract: Sexual reproduction is considered the ancestral reproductive mode of eukaryotes, yet it has been lost several times in many taxa. Understanding the mechanisms by which asexual lineages appear and persist over time remains a major challenge of evolutionary biology. During my PhD, I investigated the genetic basis as well as the evolutionary consequences of sex loss in aphids, a group that displays reproductive polymorphism. The ancestral reproductive mode of aphids is cyclical parthenogenesis (CP, an alternation of several parthenogenetic generations and one sexual generation), but obligate parthenogenesis (OP) is frequently observed in this group. Derived OP lineages are not able to produce sexual females though they often retain the ability to produce males. First, to characterize genomic regions involved in the transition from CP to OP reproductive mode, we used genome scan approaches on different aphid taxa that are more or less genetically related and exhibit variation in reproductive mode. We showed that the genetic basis of sex loss is different between the studied taxa, with no apparent convergence in gene content nor functions. Thus, several independent genomic regions may be responsible for sex loss in aphids, suggesting that there are many paths that lead to asexuality in this group. Second, we studied the evolutionary consequences of the loss of sex on traits and genes essential for sexual reproduction. Since the males produced by OP lineages are unlikely to pass on their genes (because CP lineages are usually separated from OP ones), we tested the prediction that male traits should degenerate. Male production was indeed reduced in OP lineages, supposedly resulting from counter-selection, but male reproductive success was only slightly lower than in CP lineages, presumably due to the slow action of relaxed selection or underestimation of reproductive opportunities. As OP lineages produce rare males and also do not produce sexual females, the gene expression of parthenogenetic females in these OP lineages is no longer constrained by that of other morphs. We thus predicted that the disappearance of sexual conflict (which arises when there are different morph-specific optima for a trait shared by different morphs) would result in shifts of gene expression. We therefore compared gene expression patterns of CP and OP lineages for different morphs in the pea aphid. We observed that gene expression in males from OP lineages tended towards the parthenogenetic female optimum, as predicted by theory. More surprisingly, males and parthenogenetic females of OP lineages consistently over-expressed genes typically expressed in the gonads of sexual morphs. These changes in gene expression in OP lineages may arise from the relaxation of selection or the repurposing of gene networks otherwise used in sexual lineages. This thesis illustrates the relevance of using species with polymorphic reproductive systems to understand the evolutionary history of sex loss and its consequences.

Titre : Bases génétiques et conséquences évolutives de la perte de sexe dans le groupe des pucerons

Mots clés : Parthénogénèse, transition vers l'asexualité, évolution clonale, dégénérescence de trait, expression de gènes, conflit sexuel

Résumé : La reproduction sexuée est considérée comme le mode de reproduction ancestral des eucaryotes, mais elle a été perdue à plusieurs reprises dans de nombreux taxons. Comprendre les mécanismes par lesquels les lignées asexuées apparaissent et persistent dans le temps reste un défi majeur de la biologie évolutive. Au cours de ma thèse, j'ai étudié les bases génétiques ainsi que les conséquences évolutives de la perte du sexe chez les pucerons, un groupe qui présente un polymorphisme de reproduction. Le mode de reproduction ancestral des pucerons est la parthénogenèse cyclique (CP, une alternance de plusieurs générations parthénogénétiques et d'une génération sexuée), mais la parthénogenèse obligatoire (OP) est fréquemment observée dans ce groupe. Les lignées OP dérivées ne sont pas capables de produire des femelles sexuées bien qu'elles conservent souvent la capacité de produire des mâles. Pour caractériser les régions génomiques impliquées dans la transition de la reproduction CP à OP, nous avons utilisé des approches de scan génomique sur différents taxons de pucerons plus ou moins proches génétiquement et présentant des variations de mode de reproduction. Nous avons montré que la base génétique de la perte du sexe est différente entre les taxons étudiés, sans convergence apparente dans le contenu génique ou les fonctions des gènes. Ainsi, plusieurs régions génomiques indépendantes peuvent être responsables de la perte du sexe chez les pucerons, ce qui suggère qu'il existe de nombreuses voies menant à l'asexualité dans ce groupe. Ensuite, nous avons étudié les conséquences évolutives de la perte du sexe sur les traits et les gènes essentiels à la reproduction sexuée. Comme il est peu probable que les mâles produits par les lignées OP transmettent leurs gènes (les lignées CP étant généralement séparées géographiquement des lignées OP), nous avons testé la prédiction selon laquelle les traits mâles dégénèrent. La production de mâles était en effet réduite dans les lignées OP, supposément en raison de la contre-sélection, mais le succès reproductif des mâles n'était que légèrement inférieur à celui des lignées CP, probablement en raison de la sélection relâchée qui agit lentement ou d'opportunités reproductives sous-estimées. Comme les lignées OP produisent rarement des mâles et ne produisent pas de femelles sexuées, l'expression génétique des femelles parthénogénétiques n'est plus contrainte par celle des autres morphes. Nous avons donc prédit que la disparition du conflit sexuel (qui survient lorsqu'il existe différents optima spécifiques à chaque morphe pour un trait partagé par différents morphes) entraînerait des changements d'expression des gènes. Nous avons donc comparé les profils d'expression génétique des lignées CP et OP pour différents morphes du puceron du pois. Nous avons observé que l'expression des gènes chez les mâles des lignées OP tendait vers l'optimum des femelles parthénogénétiques, comme le prévoyait la théorie. Plus surprenant, les mâles et les femelles parthénogénétiques des lignées OP surexpriment systématiquement des gènes exprimés dans les gonades des morphes sexués. Ces changements dans l'expression des gènes dans les lignées OP peuvent s'expliquer par un relâchement de la sélection ou la reconversion de réseaux de gènes autrement utilisés dans les lignées sexuées. Cette thèse illustre la pertinence de l'utilisation d'espèces aux systèmes de reproduction polymorphes pour comprendre l'histoire évolutive de la perte du sexe et ses conséquences.