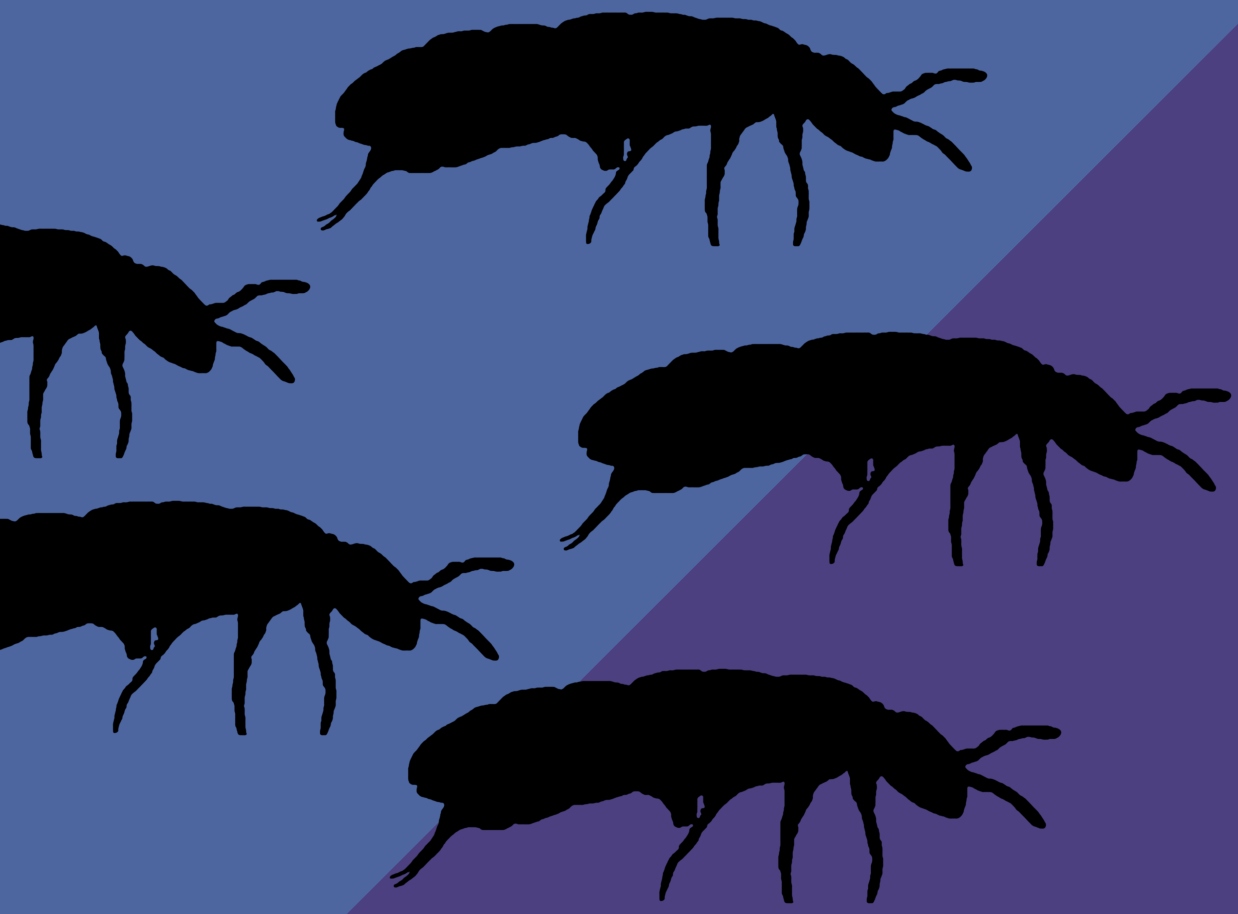
ON A FATHERLESS SPRINGTAIL AND HER LITTLE COMPANION

The fate of sexual traits and genomes under parthenogenesis



A.A. KAMPFRAATH

2026

ON A FATHERLESS SPRINGTAIL AND HER LITTLE COMPANION

The fate of sexual traits and genomes under parthenogenesis

Andries Augustus Kampfraath

On the cover and chapter artwork by Andries Augustus Kampfraath. The colour palette and visual style were inspired by Dick Bruna's cover design for Ian Fleming's *Casino Royale*. Bruna was a master at capturing the essence of a form through simplicity and carefully balanced colours. As Oscar Wild wrote in *The Philosophy of Dress*: "All good colours are equally beautiful; it is only in the question of their combination that art comes in." I tried to mimic this to the best of my abilities.

Layout: Andries Augustus Kampfraath
Printed by: ProefschriftMaken | www.proefschriftmaken.nl

ISBN: 978-94-6534-728-8
DOI: https://doi.org/10.5463/thesis.1972

© Andries Augustus Kampfraath, 2026
All rights reserved. Readers are warmly invited to build upon the ideas presented in this thesis, provided that due credit is given to their original sources. However, no part of this publication may be reproduced in any form without the written consent of the author or copyright owner.

VRIJE UNIVERSITEIT

ON A FATHERLESS SPRINGTAIL AND HER LITTLE COMPANION

The fate of sexual traits and genomes under parthenogenesis

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
volgens besluit van de decaan
van de Faculteit der Bètawetenschappen
in het openbaar te verdedigen
op woensdag 25 november 2026 om 11.45 uur
in de universiteit

door

Andries Augustus Kampfraath

geboren te Amsterdam

promotor:	prof.dr. J. Ellers
copromotor:	dr. K. Kraaijeveld
promotiecommissie:	prof.dr. M.P. Berg dr. Y. Mulla prof.dr. K. van Doninck prof.dr. E. Schranz dr. B. Visser

For Pake, who inspired me with his passion for nature,
and for my father, from whom I inherited the love for graphic design.
Their positive influences come together in these pages.

Table of contents

CHAPTER 1: General introduction 9

CHAPTER 2: Male sexual trait decay in two asexual springtail populations follows neutral mutation accumulation theory 23

CHAPTER 3: Transinfections of *Wolbachia* between *F. candida* lines; heat stress, re-infection and potential effect of bacteriophage WO 37

CHAPTER 4: Genome expansion of an obligate parthenogenesis-associated Wolbachia poses an exception to the symbiont reduction model 55

CHAPTER 5: Contrasting genome dynamics in sexual and asexual lineages of F. candida 83

CHAPTER 6: General discussion 111

References 119

List of publications 134

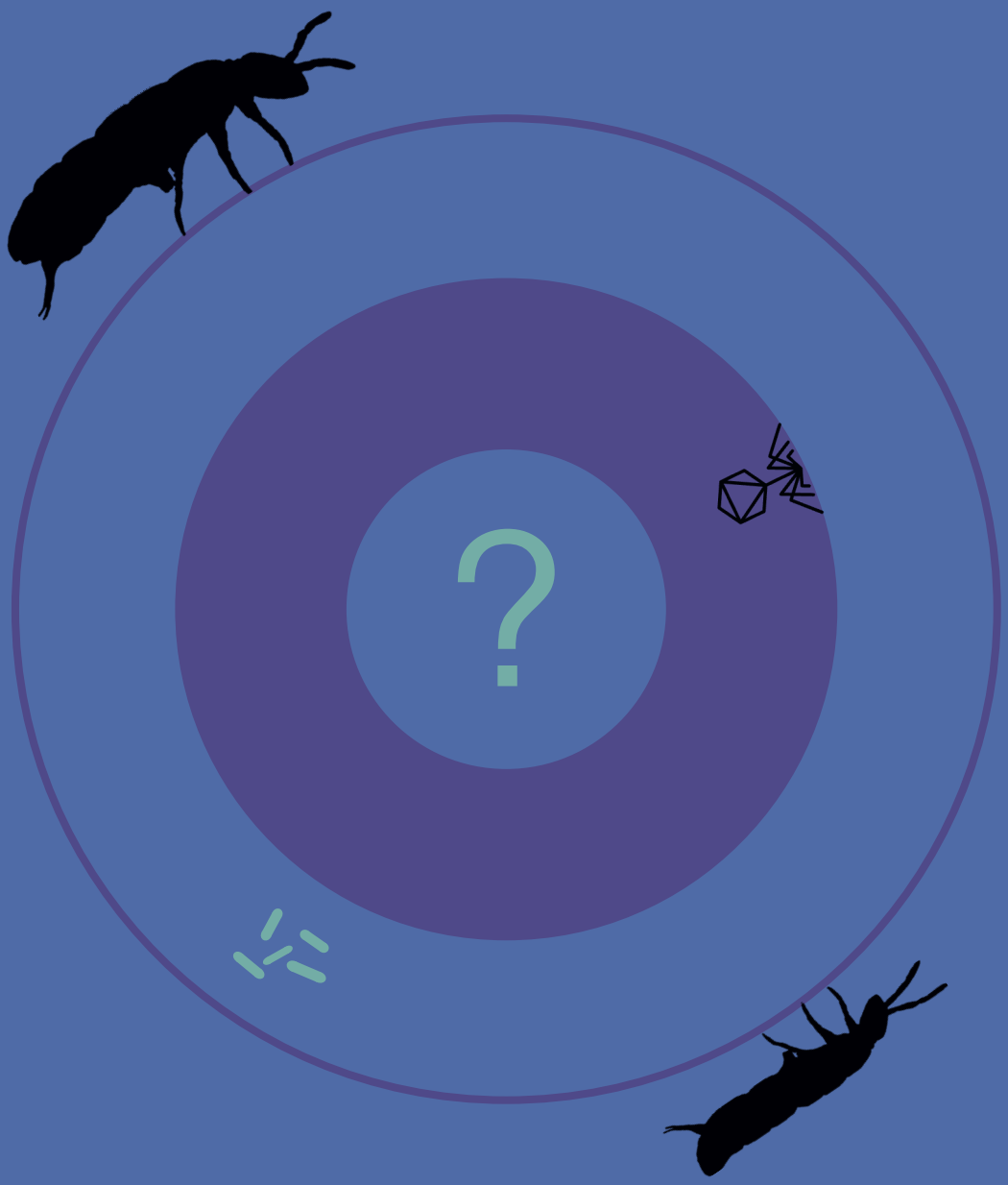
Summary 137

Samenvatting 139

Acknowledgements 143

Author 147

CHAPTER 1



DKJR

General introduction

Trait loss

When evolution is addressed, the focus is usually on the gain of traits that give a competitive advantage and can therefore lead to higher fitness. While trait gain is undoubtedly a major part of evolutionary theory, the antagonistic process on the loss of traits is often overlooked, even though trait loss can also rigorously alter the fate of a species (Ellers *et al.*, 2012). For example, the loss of energy-consuming traits that have become redundant can make organisms better competitors, as is thought to be the case for parasitoid wasps that lost the ability of synthesizing lipids (Visser and Ellers, 2008; Visser *et al.*, 2010). Loss of traits also plays an important role in the major transitions in evolution and is the focus of a growing body of literature (Fong, Kane and Culver, 1995; Lahti *et al.*, 2009; Ellers *et al.*, 2012; Womack, Fiero and Hoke, 2018). One of the most prominent examples is the loss of sleep, pigmentation and eyes in cave dwelling fish, which independently evolved multiple times across very diverse taxa, suggesting selective pressure towards losing these traits (Moran *et al.*, 2023). Another situation in which trait loss often occurs, is when a shift in the reproductive mode alters the direction and magnitude of selection pressures on the sex-specific traits of an organism (van der Kooi and Schwander, 2014).

Sexual trait loss

Sexual trait loss is likely to occur during the transition from sexual reproduction to parthenogenesis. For this form of asexual reproduction, where viable offspring develops from unfertilized eggs, sexual traits are expected to decay, as they no longer contribute to individual fitness (van der Kooi and Schwander 2014). Both male and female sexual traits are prone to decay, but at different speeds. Decay rates have been suggested to depend on whether traits are neutral (non-functional) or maladaptive (Heethoff *et al.*, 2007; van der Kooi and Schwander, 2014). For traits that are neutral within the new selective environment, the decay should result from accumulation of loss-of-function mutations, a process that is expected to be stochastic and slow (Muller, 1949; Kimura, 1983; Hall and Colegrave, 2008). By contrast, a formerly adaptive trait that is costly to maintain due to environmental conditions and/or high energy investment, is expected to deteriorate faster as a consequence of negative selection against that trait (Cooper and Lenski, 2000; Hall and Colegrave, 2008). Thus, depending on whether traits are under neutral or negative selection the rate of decay will be respectively lower or higher.
Female reproductive traits involved in mate location and copulation behaviour were found to decay faster and more consistently in asexual lineages than male

sexual traits (Heethoff *et al.*, 2007; van der Kooi and Schwander, 2014), with independent transitions to asexuality showing parallel decay of these traits (Schwander et al. 2013). However, due to the often low number of males in pathenogenetic populations, the decay of male sexual traits has only been documented for a limited number of asexual species, compared to female sexual traits (van der Kooi and Schwander, 2014). It has, for example, been found that males of three *Encarsia* endoparasitoid wasp species exhibited degraded courtship and mating behaviors despite normal sperm production, while asexually produced males of the parasitoid wasp *Muscidifurax uniraptor* and snail *Potamopyrgus antipodarum* showed sperm abnormalities (Zchori-Fein, Roush and Hunter, 1992; Giorgini, 2001; Gottlieb and Zchori-Fein, 2001; Jalinsky, Logsdon and Neiman, 2020). Decay of male traits did not follow a conserved sequence of steps across asexual lineages like female sexual traits. Male sexual traits decayed more stochastically than female sexual traits, suggesting that sexual traits of males decay as a consequence of drift rather than selective processes (van der Kooi and Schwander, 2014). However, as the number of studies of male sexual trait decay is still rather limited, the generality of these patterns is still to be determined.

Compensated trait loss in symbiotic relationships

The loss of (sexual) traits is not always apparent in the external phenotype, especially when the lost function is compensated for by species interactions. This form of trait loss is defined as compensated trait loss and often occurs in situations in which species live closely together (Ellers *et al.*, 2012). The defining feature of compensated trait loss is that a function or resource is provided by a symbiotic partner, which eliminates positive selection for the maintenance of the underlying trait. This can result in the loss of such traits and then leads to functional dependence on the partner. This type of trait loss is most commonly found in obligatory host-associated organisms and can even occur in communities of symbionts and hosts. Bacterial symbionts, for example, can have up to five to ten times smaller genomes than free-living bacteria, due to the loss of many genes. The functions of essential discarded genes were thereafter performed by their host (Figure 1). For this thesis, I will focus on two types of compensated trait loss resulting from endosymbiosis: loss of sexual traits in the host and reduction of genome size in the endosymbiont.

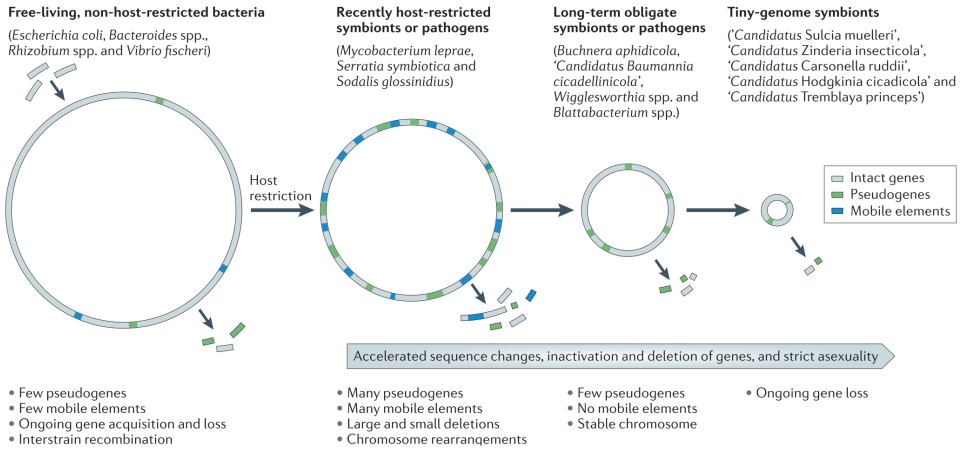


Figure 1: Stages of genome reduction in host-restricted bacteria for which small population sizes and an asexual life cycle result in mutation fixation, adapted from McCutcheon & Moran 2012.

Endosymbiosis

Species living in symbiosis have very close interactions. The most intertwined interactions are found between endosymbionts that live inside their hosts' cells. Endosymbiotic bacteria are found in the cells of many eukaryotic species, where they affect major processes such as defence against pathogens (Scarborough, Ferrari and Godfray, 2005), development (Dedeine *et al.*, 2001; Gill, Darby and Makepeace, 2014) and host reproduction (Stouthamer, Breeuwer and Hurst, 1999; Cordaux, Bouchon and Grève, 2011). Endosymbionts depend on their host for proliferation. Hosts do not necessarily depend on their endosymbiont and can often reproduce and survive in their absence. Yet, there are ample examples of obligate endosymbionts, e.g. *Buchnera* in aphids (Douglas, 1998; Shigenobu *et al.*, 2000), *Wigglesworthia* in the tsetseflies (Akman *et al.*, 2002), and *Wolbachia* in parasitic filarial nematodes (Foster *et al.*, 2005; Darby *et al.*, 2012). In these cases, neither host nor endosymbiont are viable without the other, and these associations are usually characterised by a long evolutionary history, nutritional or developmental dependency, and vertical transmission of the symbiont (Moran and Bennett, 2014). Obligate endosymbionts, like *Wolbachia* and *Cardinium*, are even capable of inducing asexuality in their host, which benefits their spread through the population as they are vertically transmitted from mother to their offspring (Zchori-Fein *et al.*, 2001; Werren, Baldo and Clark, 2008). All these complex interactions strongly shape the genome of the endosymbiont and have played an important role in major evolutionary events like the development of the eukaryotic cell (McCutcheon and Moran, 2012; Archibald, 2015; Keeling and McCutcheon, 2017).

Genomic consequences of endosymbiosis

Current theory predicts that mutual dependency between host and endosymbiont leads to a reduction in genome size of endosymbionts (Figure 1)(Moran, McCutcheon and Nakabachi, 2008; McCutcheon and Moran, 2012). Different genetic mechanisms have been implied as the driving forces behind these reductions. First, co-adaptation between the symbiont and their host may result in redundancy of certain symbiont functions releasing genes from selective constraints, causing them to decay and eventually to disappear (Dale and Moran, 2006; Ellers *et al.*, 2012). Second, small population sizes and limited opportunity for horizontal gene exchange increases genetic drift to a level that purifying selection cannot overcome. This leads to deleterious mutations that cause pseudogenization of mildly advantageous genes and the eventual removal of those genes due to the inherent deletion bias found for bacterial genomes (Moran, 1996; Mira, Ochman and Moran, 2001; McCutcheon and Moran, 2012). In addition, multiple transitional events, such as becoming host bound, moving into a specialized host cell, and being vertically transmitted through host generations, seem to pinpoint moments of severe genome reduction (Lo, Huang and Kuo, 2016). Intervals between different transitional events are subject to varying selection pressures that can affect genome size. For example, the negative correlation between host dependence and symbiont genome size only holds for vertically transmitted endosymbionts (Fisher *et al.*, 2017). Nonetheless, the genome size of vertically transmitted endosymbionts not residing in specialized host cells varies a lot, which is difficult to explain in the context of current theory (Lo, Huang and Kuo, 2016). Examining endosymbiont genomes in this stage will be needed to better understand all factors that influence the genome reduction during the evolution of endosymbionts.

Genomic consequences of asexuality

In addition to the genomic consequences seen in endosymbionts, there are also expected changes in the genomes of species that changed to an asexual lifestyle. This major lifestyle change can affect a great variety of genomic traits, e.g. mutation accumulation, adaptive evolution, homozygosity, palindromes, gene conversion, increased numbers of transposable elements, horizontal gene transfers, gene family expansions and gene loss (Figure 2) (Normark, Judson and Moran, 2003; Kraaijeveld *et al.*, 2012, 2016; Ament-Velásquez *et al.*, 2016; Tvedte, Logsdon and Forbes, 2019; Jaron *et al.*, 2020). However, it has been difficult to draw general conclusions about the genetic consequences of transitioning to asexual reproduction, as the different studies show inconsistent changes in the presence or absence of specific genetic elements (Jaron *et al.*, 2020; Tvedte, Logsdon and Forbes, 2019). A focus on overall larger scale genome dynamics instead of specific

genetic elements might give us better insight into the genomic changes in asexual species.

Major improvements in the availability and affordability of long-read sequencing techniques have facilitated the formation of highly contiguous genome assemblies that enable us to shift the focus in evolutionary genomics from gene content to genome structure and dynamics. This resulted in several studies that reported the occurrence of large-scale changes in the form of segmental duplications within the genomes of asexual organisms e.g. rotifers, nematodes and springtails (Flot *et al.*, 2013; Blanc-Mathieu *et al.*, 2017; Faddeeva-Vakhrusheva *et al.*, 2017; Nowell *et al.*, 2018). Strikingly, these studies all found large palindromes, a special kind of segmental duplication potentially common in asexual genomes. Palindromes are duplicated regions on a single chromosome, that can align and form hairpins because of their reverse orientation. This mechanism allows for gene conversion within duplicated regions and has been shown to purge mutations from Y-chromosomes in humans and chimpanzees (Rozen *et al.*, 2003; Trombetta and Cruciani, 2017). The formation of hairpins could also hinder chromosome pairing during meiosis, which is the reason why palindromic regions are thought to be an indication of non-recombining areas within the genome, like Y-chromosomes and chromosomes of ameiotic lineages. However, palindromes do not seem to be consistently present in all studied asexual genomes (Jaron *et al.*, 2020), this might partly depend on the quality and contiguousness of the genome assemblies or the time they take to evolve. Therefore, the relationship between the presence of palindromic regions and asexuality is still unresolved, leaving open the possibility that palindromic presence is a species-specific trait rather than associated with asexuality. Comparing more genomes of paired sexual and asexual linages could provide us with further knowledge on whether palindromes within non-sex chromosomes are signs of asexuality.

Research questions

My thesis will focus on answering the broad research question whether a switch from sexual reproduction to asexual reproduction under influence of an endosymbiont results in compensated trait loss. More specifically, I will address the questions:

- Are male sexual traits affected by the switch from sexual to asexual reproduction?
- Is the switch to asexual reproduction reversable and reproduceable?
- What genomic effects will the shift in reproductive system of the host have on endosymbiotic partners?
- What genomic features are characteristics of the switch from sexual to asexual reproduction?

Below, the study system will be introduced and more concrete questions that follow from these broader research questions will be presented.

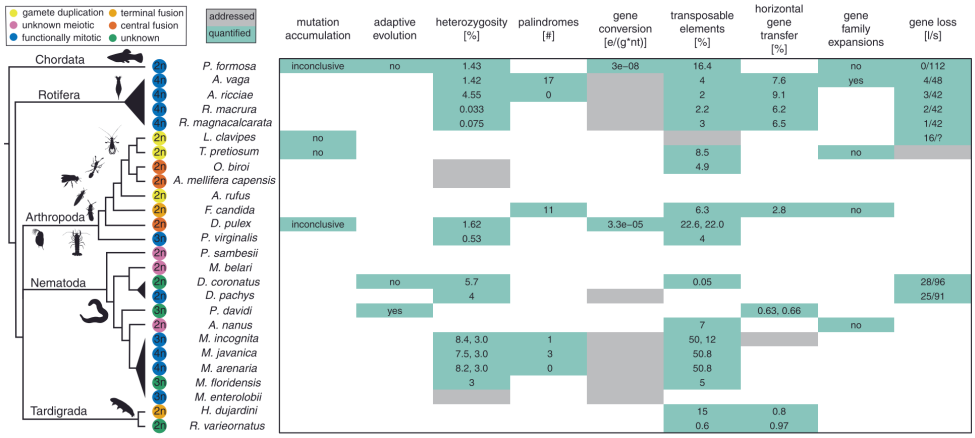


Figure 2: Genome features studied in parthenogenetic animal species, adapted from Jaron et al., 2020. The phylogeny displays the taxonomic relationships of the 26 sequenced parthenogenetic animal species considered here, representing at least 18 independent transitions to parthenogenesis from 5 different animal phyla. Species that might derive from the same original transition are grouped in triangles. The colour of the circle indicates the cellular mechanism of parthenogenesis and the number inside the circle the ploidy of the species. Green cells represent cases where quantified information of the genomic feature was available in literature (values are indicated); the grey cells represent cases where the genomic features were addressed with respect to parthenogenesis, but the results were not quantitatively comparable to other studies. Findings for mutation accumulation and adaptive evolution refer to comparisons between sexual and asexual species and are reported with respect to theoretical predictions (yes: as predicted, no: opposite to predictions, inconclusive: no difference). e/(g*nt): **e**vent per **g**eneration per **n**ucleo**t**ide; l/s: number of **l**ost genes among the **s**tudied genes related to sexual reproduction.

Study system

The study system in which these broader research questions will be explored, is the well-known springtail *Folsomia candida* (Box 1) and its endosymbiont *Wolbachia* (Box 2), specifically the *w*Fol strain.

Folsomia candida

F. candida is a member of the order Collembola, informally known as springtails. It is a common and widespread arthropod that occurs in soils throughout the world and has been used as a "standard" test organism for ecotoxicological experiments

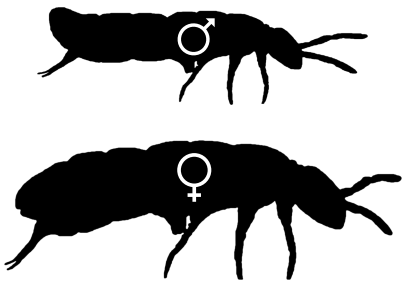
(Fountain and Hopkin, 2005). Both sexual and asexual lineages have been described (Goto, 1960; Fountain and Hopkin, 2005; Tully *et al.*, 2006; Krogh *et al.*, 2008; Tully and Potapov, 2015); (Goto, 1960; Frati *et al.*, 2004; Buono and Tully, 2011). Asexual lines are the most widely studied and carry the obligate intracellular bacterium *Wolbachia. Wolbachia* plays a crucial role for successful parthenogenetic development of females, as antibiotic treatment inhibits egg development in *F. candida* (Pike and Kingcombe, 2009; Timmermans and Ellers, 2009). In line with what is observed in a variety of animal taxa, asexual females occasionally produce males (van der Kooi and Schwander, 2014). *F. candida* is a diplo-diploid species (XO:XX), with males being the heterogametic sex. Although the exact cellular mechanism underling the rare male development is not confirmed, it can be hypothesized that an error occurs during the fusion of the haploid egg with the polar body or during the post-zygotic development, leading to the elimination of an X chromosome. Males are produced at a very low rate by asexual *F. candida*, with only one male out of 1000 to 10000 individuals (Krogh *et al.*, 2008; Buono and Tully, 2011). These rare males seem to be sterile, as keeping them in an equal ratio together with females does not lead to shared offspring, nor are they able to fertilise the eggs of sexual populations (Krogh *et al.*, 2008; Buono and Tully, 2011). Thus, male trait decay is expected to have occurred in asexual *F. candida*. However, which sexual traits are degraded and to what extent has yet to be tested.

The life cycle and reproductive behaviour of the less studied sexual lines have not been fully described. Evidence suggests that sexual individuals of *F. candida* have an equal sex ratio (Goto, 1960; Frati *et al.*, 2004; Buono and Tully, 2011) and transfer their sperm indirectly via sperm packages (spermatophores) deposited on the soil. A common mode of reproduction in soil arthropods is where females collect a spermatophore for internal fertilisation without any physical contact with a male (Schaller, 1971; Proctor, 1998; Frati *et al.*, 2004; Buono and Tully, 2011; Walter and Proctor, 2013; Zizzari, Jessen and Koene, 2017). In such a mating system, referred to as dissociated sperm transfer, spermatophore-related traits play a crucial role in male reproductive success, as the spermatophores convey all the information used by females to assess the quality of their 'mate' (Zizzari *et al.*, 2009; Zizzari, van Straalen and Ellers, 2013). Thus, assessment of the life cycle and reproductive behaviour of a sexual line is needed to get insight into the degradation of these traits.

F. candida genome

In 2017 the first high-quality genome of the asexual Berlin line of *F. candida* was published (Faddeeva-Vakhrusheva *et al.*, 2017), in which multiple unexpected features were found. This included an expansion of stress response gene families, many HGT genes related to lignocellulose degradation and changes in the Hox gene cluster arrangement. Most remarkable were the occurrences of intra- as well as

inter-scaffold large-scale duplications. Some of these formed palindromes that, as described above, could be a sign of asexual reproduction. Eleven to fifteen of these palindromes were detected that co-occurred with transposable elements (TE) and horizontally transferred genes, indicating that they are in regions with elevated genome dynamics. Recently, a sexual and asexual line of *F. candida* were sequenced and while differences in genome length and number of transposable elements were found, palindromes were not examined (Luan *et al.*, 2023). Comparing genomes of sexual and asexual linages could provide us with the answer to the question whether palindromes within non-sex chromosomes are signs of asexuality.

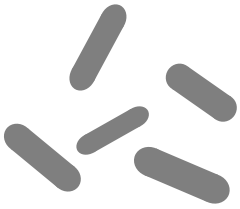


Name: ***Folsomia candida***
Willem 1902

Taxonomic classification

Domain:	Eukaryota
Kingdom:	Animalia
Phylum:	Arthropoda
Class:	Collembola
Order:	Entomobryomorpha
Family:	Isotomidae
Genus:	*Folsomia*
Species:	*F. candida*
Size:	1.5-3 mm

Box 1: *F. candida* classification



Name: ***Wolbachia***
Hertig 1936

Taxonomic classification

Domain:	Bacteria
Phylum:	Pseudomonadota
Class:	Alphaproteobacteria
Order:	Rickettsiales
Family:	Ehrlichiaceae
Genus:	*Wolbachia*
Strain:	*w*Fol
Gram stain:	negative
Size:	0.8-1.5 µm

Box 2: *Wolbachia* classification

Wolbachia

Wolbachia is one of the most common and widespread bacterial endosymbionts, that is found in a wide range of invertebrate and nematode species and is estimated to potentially infect 66% of all arthropod species (Hilgenboecker *et al.*, 2008; Zug and Hammerstein, 2012). It is a gram-negative, α-proteobacteria from the

Rickettsiales order, best known for its variety of interactions with its host, that differ strongly in their nature. In nematodes they are mutualistic and, in some cases, provide means necessary to function properly, while in arthropods they interfere with the reproductive system (Werren, Baldo and Clark, 2008; Kent and Bordenstein, 2010). These reproductive manipulations include: **feminization**, where genetic males develop as females, **male killing**, where males are feminized and die during larval development, **induction of parthenogenesis** (PI), causing unfertilized eggs to develop into females and **cytoplasmic incompatibility** (CI) in which sperm from *Wolbachia*- infected males is not compatible with eggs from females uninfected with the same strain (Figure 3)(Werren, 1997; Werren, Baldo and Clark, 2008; Merçot and Poinsot, 2009). Utilizing the combined effects of CI, antiviral capabilities and host fitness reductions in mosquitoes has led to a whole new field of research on combatting the spread of viral diseases, under which Dengue and Zika by infecting midges with selected Wolbachia strains. The release of transinfected mosquitoes that can invade and replace the native populations has shown to be an effective tool for disease control under the right conditions (Ross, 2024).

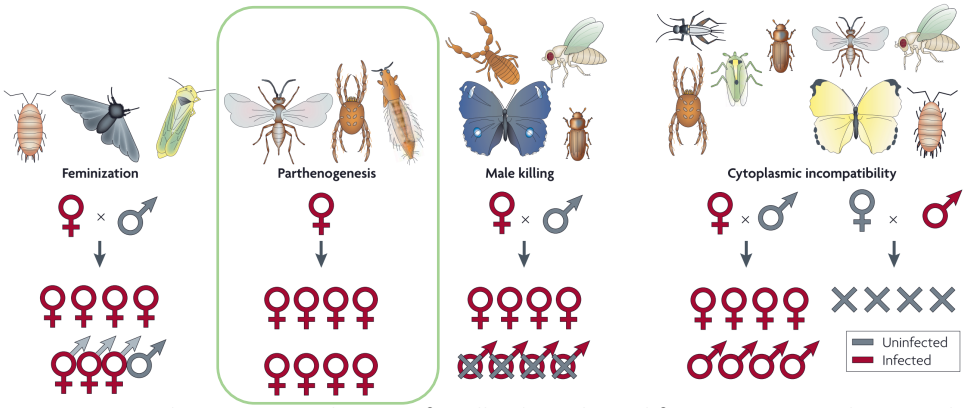


Figure 3: Reproductive manipulations of Wolbachia adapted from Werren et al. 2008. The green box highlights parthenogenesis, which is linked to *Wolbachia* in *F. candida*

Transmission & Transinfection

The main mode of *Wolbachia* transmission is to spread vertically from mother to offspring (Werren, Baldo and Clark, 2008). However, a lack of concordance between the phylogeny of *Wolbachia* and the phylogeny of the hosts, indicates that there is extensive horizontal transfer of *Wolbachia* between hosts species (Vavre *et al.*, 1999). This can occur, for example, through predation, blood contact, host-parasitoid interactions, or through sharing of the same host egg in parasitoid wasps (Rigaud and Juchault, 1995; Vavre *et al.*, 1999; Huigens *et al.*, 2004; Le Clec'h,

Raimond, *et al.*, 2013). In the lab, artificial horizontal transfer of *Wolbachia* by transinfection has been used to, control arthropod-borne diseases and investigate *Wolbachia*-host interactions (Hughes and Rasgon, 2014). It is both possible to inject the eggs or adults, with the ultimate objective being to infect the germ line to create a newly infected line. However, transinfection attempts are often unsuccessful for reasons unknown.

Wolbachia genetics and genomics

Wolbachia is a genetically highly diverse genus that is now divided into at least 19 supergroups (monophyletic clusters), named with letters A to U, excluding G and R (Olanratmanee *et al.*, 2021; Mioduchowska *et al.*, 2023) and possibly 21 (Sharma and Som, 2023). All known *Wolbachia* strains are in a similar transitional stage from free-living to obligate endosymbiont bacteria according to the classification of Lo et al. (2016), i.e. as they are mainly transmitted vertically and do not reside in specialized structures. They are thus predicted to vary in genome size depending on the host dependency (Lo, Huang and Kuo, 2016). This prediction seems to hold for most of the *Wolbachia* strains, with reduced genomes that lack mobile elements found in obligate strains and larger genomes in facultative strains (Brown *et al.*, 2016). However, recently a *Wolbachia* genome was sequenced that poses an exception to this pattern. The *Wolbachia* from the parthenogenetic collembolan host *Folsomia candida* (*w*Fol) has the largest *Wolbachia* genome sequenced so far (Faddeeva-Vakhrusheva *et al.*, 2017), yet it is obligate to its host. In addition, the *w*Fol genome is the first fully sequenced genome from supergroup E that is positioned to be a sister group to supergroup A, B, C, D, F and H, and shares a more ancestral common ancestor with supergroup L (Figure 4) (Gerth *et al.*, 2014; Brown *et al.*, 2016). All complete genomes from members of these supergroups have smaller genomes than *w*Fol, which suggests that this obligate *Wolbachia* strain showed genome expansion rather than the reduction that would be predicted by the symbiont genome reduction model. Thus, comparing the *w*Fol to other *Wolbachia* could shed better light on the genome evolution of obligate endosymbionts and the presumed trait loss.

Furthermore, the *w*Fol genome could contribute to understanding the genes and mechanisms underlying PI, as *w*Fol has been suggested to cause parthenogenesis in its host. All parthenogenetic *F. candida* lines carry *Wolbachia*, whereas the sexually reproducing lines do not (Czarnetzki and Tebbe, 2004; Frati *et al.*, 2004; Timmermans and Ellers, 2009). This is remarkable as PI producing *Wolbachia* have so far mostly been found in haplo-diploid species, while *F. candida* is a diploid-diploid species (Frati *et al.*, 2004). However, unlike haplo-diploid species where the link between *Wolbachia* infection and PI is unequivocal (Russell and Stouthamer, 2011), in *F. candida* the exact role of *w*Fol in parthenogenesis is not yet fully understood. *w*Fol has been observed around cytoplasmic asters in the eggs

of *F. candida* during meiosis (Riparbelli, Giordano and Callaini, 2006) and elimination of *Wolbachia* by heat or antibiotics, renders the host's eggs non-viable (Pike and Kingcombe, 2009; Timmermans and Ellers, 2009). However, it cannot be completely ruled out that the lack of egg viability is caused by the curing treatments rather than the absence of *Wolbachia*. Moreover, due to the infertility of cured females no cured lines can be established to test this. Thus, the question remains whether Wolbachia is the cause of PI in *F. candida*. This could be unequivocally proven by the establishment of a new asexual line from transinfecting a sexual line with *w*Fol.

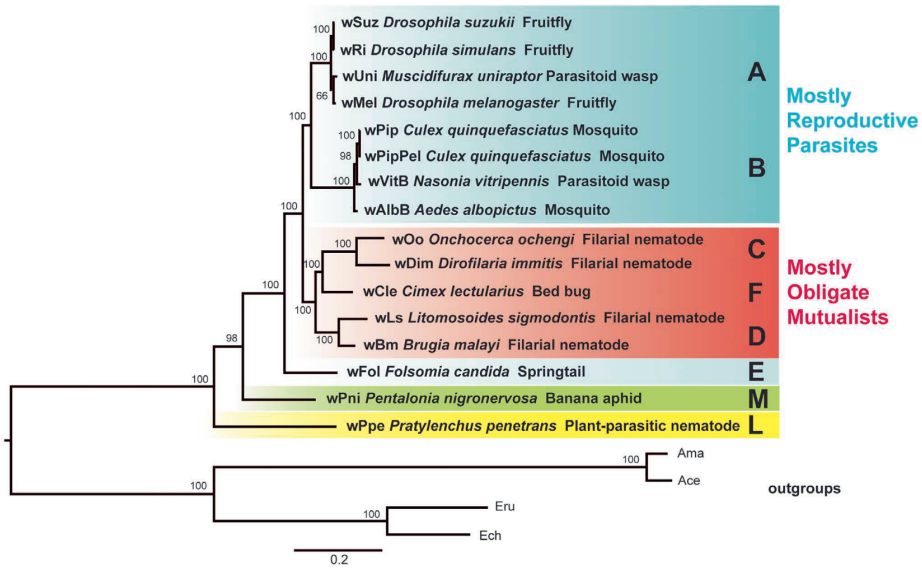


Figure 4: Maximum likelihood phylogeny of Wolbachia supergroups A, B, C, D, E, M and L, based on 79 conserved single-copy orthologous genes adapted from Brown et al. 2016. The tree was generated based on 61,465 nucleotide alignment positions with RAxML under the GTR model. Bootstrap values shown on branches are from 1,000 replicates. Shading indicates major groups.

Phage WO

A further complication in the interaction between *F. candida* and *w*Fol is the presence of bacteriophages (Kampfraath *et al.*, 2019; Bordenstein and Bordenstein, 2022). A bacteriophage is a type of virus that specifically infects bacteria. There are two types of bacteriophages: virulent and temperate bacteriophages. Virulent bacteriophages only reproduce by means of a lytic cycle, in which they insert their genetic material into the bacteria and use the cell's machinery to produce new bacteriophages that eventually lyse the bacterial cell and releases the

bacteriophage particles. Temperate bacteriophages do not only have a lytic cycle but also a lysogenic cycle, in which the bacteriophage genetic material gets integrated in the DNA of the bacterial host, called a prophage (Campbell 2003). Prophages have been found in many other *Wolbachia* (Bordenstein and Bordenstein, 2022) and within wasps there is evidence that phage and bacterial densities are inversely related, suggesting that these prophages go into the lytic cycle and thereby kill *Wolbachia* cells (Bordenstein *et al.*, 2006). The lytic cycle can be triggered when the host of the phage encounters stress, such as temperature and host age (Bordenstein and Bordenstein 2011, Lu et al. 2011). For the prophages in *w*Fol it is still unknown whether they can still become lytic and if so under what conditions.

Outline of the thesis

This thesis will address the stated research questions in the context of the study system of the soil dwelling springtail *F. candida* and its *Wolbachia* endosymbiont *w*Fol.

In *Chapter 1: Male sexual trait decay in two asexual springtail populations follows neutral mutation accumulation theory*, we investigated male sexual traits in one sexual and two parthenogenetic populations of *F. candida.* Our study provides the first detailed description of the reproductive behaviour of sexual *F. candida.* We analysed several spermatophore-related traits: stalk length, sperm droplet size, amount of sperm DNA material and attractiveness. Specifically, we tested whether these traits are subject to decay and whether a parallel trait regression occurred in the two asexual *F. candida* lines.

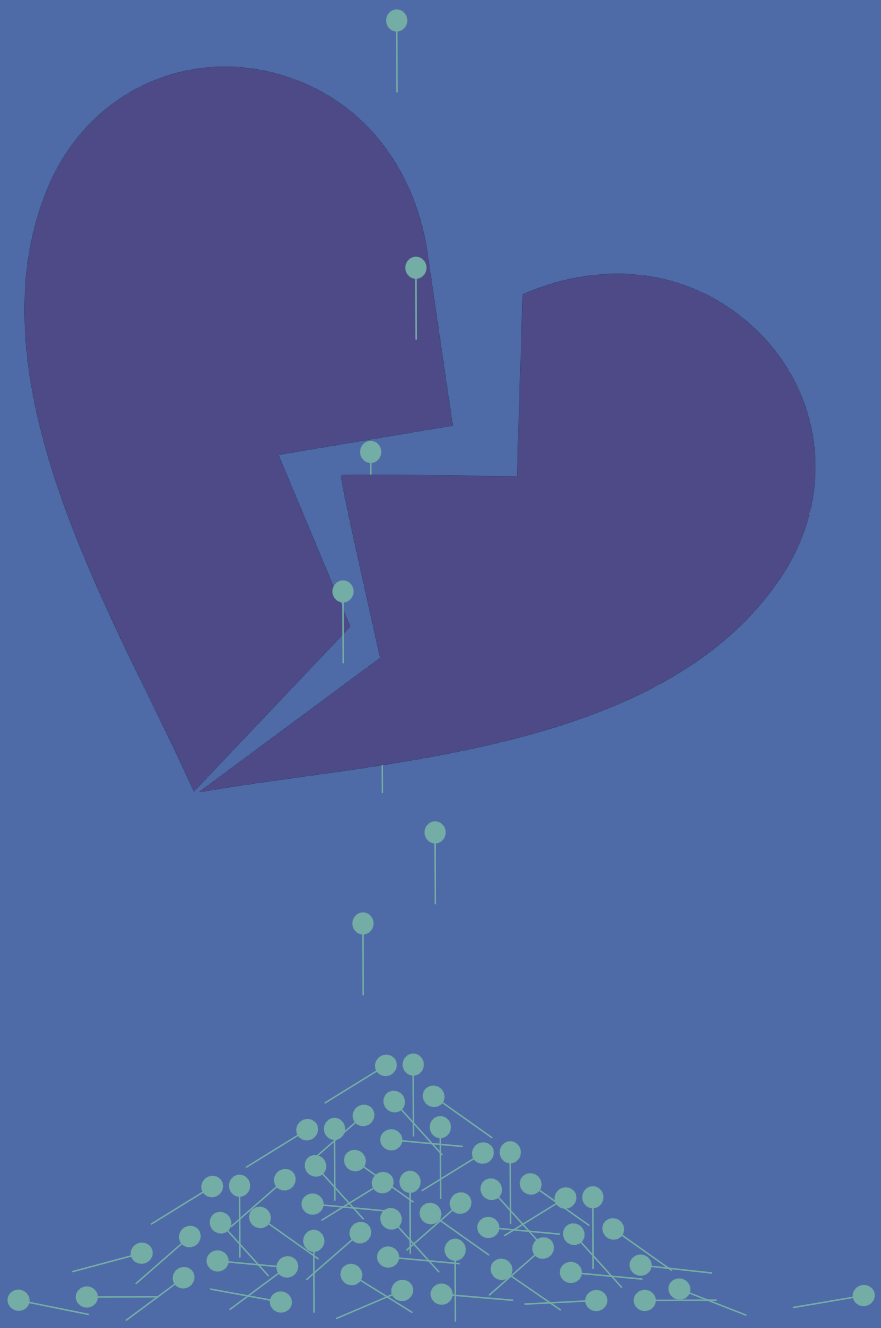
In *Chapter 2: Transinfections of wFol between F. candida lines; re-infection and potential effect of bacteriophage WO*, we attempted to create a newly infected line of *F. candida* by transferring *w*Fol into the females of the sexual line. As the infection did not establish, we attempted to determine whether *Wolbachia* could be exterminated in the newly infected hosts by activation of lytic cycle of the phages within its bacterial genome. To address this question, we measured the amount of *Wolbachia* and phage WO after transfer from the parthenogenic line into the sexual line. Activation of phages by transinfection has not been reported before, therefore we also examined whether heat stress treatment can cause the phages within wFol to go into the lytic cycle and affect the relative abundances of *Wolbachia* and phage WO. In addition, we tested if curing and reinfection of the parthenogenetic line influences egg hatching to gain stronger evidence of *Wolbachia*'s role in the parthenogenetic reproduction of *Folsomia candida.*

In *Chapter 3: Genome expansion of an obligate parthenogenesis-associated Wolbachia poses an exception to the symbiont reduction model*, we set out to explain the factors that contribute to the expansion in the *w*Fol genome, which can help further understand the genomic evolution within endosymbionts. Therefore, we updated and re-annotated the *w*Fol genome and compared it to a diverse set of high quality *Wolbachia* genomes. In addition, we searched for genes shared by *Wolbachia* associated with PI that could aid in understanding the mechanism behind this manipulation. To this end, we assembled a draft genome of an additional PI strain, *w*Lcla, that has been shown to cause diploidy restoration in the parasitoid wasp *Leptopilina clavipes* (Pannebakker *et al.*, 2004). Finding the genes underlying PI in *Wolbachia* might elucidate the mechanism behind it and could resolve the debate on the nature of *w*Fol host interaction.

In *Chapter 4: Comparing genome dynamics between sexual and asexual lines of F. candida*, we focussed on the question whether the large-scale genomic changes, including palindromes, are caused by the parthenogenetic lifestyle. We sequenced, assembled and annotated the genomes of two asexual and one sexual line of *F. candida* and screened them for segmental duplications and other repetitive genomic elements. We determined whether genome dynamics differed between the sexual and asexual lines by looking into the changes of these elements over time. Furthermore, we investigated if the palindromes found in *F. candida* are linked to asexuality or if they are a species-specific trait.

Finally, in the general discussion, we reflect upon the combined results of the research chapters and put them into a broader context.

CHAPTER 2



DKJR

Male sexual trait decay in two asexual springtail populations follows neutral mutation accumulation theory

Andries Augustus Kampfraath, Tjeerd Pieter Dudink, Ken Kraaijeveld, Jacintha Ellers & Zaira Valentina Zizzari

Status: published in Evolutionary Biology

Abstract: The transition to asexual reproduction is frequent and widespread across the tree of life and constitutes a major lifestyle change. Without sexual reproduction, selection on sexually selected traits is expected to be weaker or absent, allowing the decay and ultimately loss of sexual traits. In this study, we applied an experimental approach to investigate the decay of reproductive traits under asexuality in two asexual populations of the springtail *Folsomia candida*. Specifically, we compared several key male sexual traits of a sexual population and two distinct parthenogenetic lines. To allow direct comparisons between sexual and asexual individuals we first determined the life history characteristics of the sexual *F. candida*, which perform an indirect transfer of sperm packages (spermatophores). To investigate the decay of male sexual traits under asexuality we measured the size of spermatophores, quantified the amount of sperm DNA material within spermatophores, and tested spermatophore attractiveness to females in all three populations. The amount of sperm in the sperm droplet and the attractiveness of spermatophores were lower in the asexual lines compared to the sexual population. However, the two asexual lines differed in the extent of decay of these traits. Our results are consistent with predictions from neutral mutation accumulation theory, and thus suggest this to be the main evolutionary process underlying the decay of male traits in *F. candida*.

Keywords: Collembola, spermatophore, indirect sperm transfer, trait decay, asexuality, parthenogenesis

Introduction

Loss of traits plays an important role in the major transitions in evolution and is the focus of a large body of literature (Fong, Kane and Culver, 1995; Lahti *et al.*, 2009; Ellers *et al.*, 2012). A shift in the reproductive mode of an organism (e.g., loss of sexual reproduction) can alter the direction and magnitude of selection pressures on traits that no longer contribute to individual fitness, such as sex-specific traits. In the transition from asexuality to parthenogenesis, a form of asexual reproduction that gives raise to offspring from unfertilized eggs, sexual traits are expected to decay (van der Kooi and Schwander 2014). Both male and female sexual traits are subject to decay. However, the decay of male sexual traits has been documented for a limited number of asexual species, compared to female sexual traits (van der Kooi and Schwander, 2014). For example, males of three endoparasitoid wasp species of the *Encarsia* genus showed degradation of courtship and mating behaviours, although the sperm production was still normal (Zchori-Fein, Roush and Hunter, 1992; Giorgini, 2001). Evidence for sperm abnormalities in asexually produced males was found in the parasitoid wasp *Muscidifurax uniraptor* (Oliver, Tanaka and Sawada, 1973; Gottlieb and Zchori-Fein, 2001) and the snail *Potamopyrgus antipodarum* (Oliver, Tanaka and Sawada, 1973; Gottlieb and Zchori-Fein, 2001).

Interestingly, some female traits such as the ones involved in mate location and copulation behaviours have been shown to decay faster compared to male sexual traits (Heethoff *et al.*, 2007; van der Kooi and Schwander, 2014). A difference in rate of decay has been suggested to depend on whether these traits are neutral (non-functional) or maladaptive (Heethoff *et al.*, 2007; van der Kooi and Schwander, 2014). For traits that are neutral within the new selective environment, the decay should result from accumulation of loss-of-function mutations, a process that is expected to be stochastic and slow (Muller, 1949; Kimura, 1983; Hall and Colegrave, 2008). On the contrary, a formerly adaptive trait costly to maintain due to environmental selection (i.e. predation) and/or to high energy investment, is expected to deteriorate faster as a consequence of negative selection against that trait (Cooper and Lenski, 2000; Hall and Colegrave, 2008).

Female reproductive traits were found to decay faster and more consistently in the majority of the measured asexual lineages (van der Kooi and Schwander, 2014), with independent transitions to asexuality showing parallel decay of these traits (Schwander et al. 2013). Conversely, no evidence of parallel male trait decay was found across asexual lineages. This suggests that sexual traits of males decay more slowly, as a consequence of drift rather than selective processes (van der Kooi and Schwander, 2014).

In the springtail *Folsomia candida* Willem, both sexual and asexual lineages have been described (Goto, 1960; Fountain and Hopkin, 2005; Tully *et al.*, 2006; Krogh *et al.*, 2008; Tully and Potapov, 2015); (Goto, 1960; Frati *et al.*, 2004; Buono

and Tully, 2011). Asexual lines carry the obligate intracellular bacterium *Wolbachia,* who plays a crucial role for successful parthenogenetic development of females, as antibiotic treatment inhibits egg development in *F. candida* (Pike and Kingcombe, 2009; Timmermans and Ellers, 2009) Pike and Kingcombe 2009). In line with what is observed in a variety of animal taxa, asexual females occasionally produce males (van der Kooi and Schwander, 2014). *F. candida* is a diplodiploid species (XO:XX), with males being the heterogametic sex. Although the exact cellular mechanism underling the rare male development is not confirmed, it can be hypothesized that an error occurs during the fusion of the haploid egg with the polar body or during the post-zygotic development, leading to the elimination of an X chromosome. Males are produced at a very low rate by asexual *F. candida*, with only one male out of 1000 to 10000 individuals (Krogh *et al.*, 2008; Buono and Tully, 2011). These rare males seem to be sterile, as keeping them in an equal ratio together with females does not lead to shared offspring, nor are they able to fertilise the eggs of sexual populations (Krogh *et al.*, 2008; Buono and Tully, 2011). Thus, male trait decay is expected to have occurred in asexual *F. candida*. However, which sexual traits are degraded and to what extent has yet to be tested.

In this study, we investigated male sexual traits in one sexual and two parthenogenetic populations of *F. candida.* Although *Folsomia candida is a* widely established model organism, the reproductive behaviour of males and females has not been fully described. Evidence suggests that sexual individuals of *F. candida* have an equal sex ratio (Goto, 1960; Frati *et al.*, 2004; Buono and Tully, 2011) and transfer their sperm indirectly via sperm package (from now referred to as spermatophore) deposited on the soil. However, it is unknown whether the reproductive biology of sexual *F. candida* is in line with what is observed in other collembolans. Our study provides the first detailed description of the reproductive behaviour of sexual *F. candida.*

A common mode of reproduction in soil arthropods, consists of females picking up a spermatophore for internal fertilisation without any physical contact with the male (Schaller, 1971; Proctor, 1998; Frati *et al.*, 2004; Buono and Tully, 2011; Walter and Proctor, 2013; Zizzari, Jessen and Koene, 2017). In such a mating system, referred to as dissociated sperm transfer Alexander (1964), spermatophore-related traits play a crucial role in male reproductive success, as the spermatophores convey all the information used by females to assess the quality of their 'mate' (Zizzari et al. 2009, 2013). Therefore, in this study we analysed several spermatophore-related traits: stalk length, sperm droplet size, amount of sperm DNA material and attractiveness. Specifically, we tested whether these traits are subject to decay and whether a parallel trait regression occurred in both asexual *F. candida* lines.

Methods

Folsomia candida cultures

One sexual and two parthenogenetic lines of *F. candida* were used. The sexual population (from now on "CR") was collected in Cremona (Italy) and described by (Frati *et al.*, 2004). One of the parthenogenetic lines used was the "Berlin strain" (from now on "BE") which is commonly used for lab tests (Vrije Universiteit, Amsterdam). The other parthenogenetic line used in this study was collected in a ground water sample in October 2013 in Evergem (from now on "EV"), Belgium, and kept in the lab of Vrije Universiteit Amsterdam since. All cultures were kept in pots with a water-saturated plaster of Paris bottom containing 10% charcoal at 20 °C in a 16:8 light dark regime and were fed dried baker's yeast (Dr. Oetker) *ad libitum*.

Reproductive biology of sexual F. candida

The reproductive biology of several collembolan species has been described in detail. Upon maturation males and females alternate reproductive and non-reproductive instars, separated by moults (Poggendorf 1956; Waldorf 1971, 1975, Zizzari et al. 2009, 2017a). Males deposit the spermatophores only during reproductive instars, while females pick up a single spermatophore at the onset of each reproductive instar to fertilise all the eggs, as they do not store sperm (Waldorf, 1971; Gols, Ernsting and Van Straalen, 2004; Dallai, Zizzari and Fanciulli, 2008). To characterize the reproductive biology of the sexual CR population, collembolans were kept individually in plexiglas vials (2.4 cm) with a bottom of water-saturated plaster of Paris containing 10% charcoal and a lid with fine meshed gauze at culture conditions. To measure the juvenile instar length (in days) and determine age of first reproduction, hatchlings were isolated (N=20) and checked daily for the presence of moulting skins. Females will only start laying eggs after taking up a spermatophore(Zizzari *et al.*, 2009; Zizzari and Ellers, 2014; Zizzari, van Straalen and Ellers, 2016), therefore, nearly adult females were offered freshly laid spermatophores soon after moulting, and were subsequently checked for egg laying, following a protocol established for other collembolans (Zizzari et al. 2009, 2013; Zizzari and Ellers 2014). As for the males, we monitored each individual for the presence of spermatophores in the vial. After reaching reproductive maturity, the number of spermatophores produced daily (males, N=30) and the number of eggs (females, N=20) were scored during their first reproductive instar. For first egg clutches (N=20), clutch size, hatching time and sex ratio were also determined. During the characterisation of the reproductive biology, we noticed that male reproductive instars lasted longer than non-reproductive instars, thus we tested if this difference was significant by comparing the lengths of the first reproductive

instars with the lengths of the subsequent non-reproductive instar. Finally, we tested if female reproductive instar length would change when no spermatophore was offered by comparing instar lengths of females in the 10th instar with and without spermatophores present (N=24).

Male sexual traits

Individually kept males of the three lineages (CR, N=32 ; BE, N=23 ; EV, N=9) were daily monitored for the presence of spermatophores to be used for the measurement of stalk length, sperm droplet size, amount of sperm DNA material, and for conducting the female choice trials.

Morphological characteristics of the spermatophore

Spermatophores of the sexual (N=21) and asexual lines (EV, N=9; BE, N=12) were collected and mounted on microscope slides to measure stalk length, sperm droplet size, and amount of sperm DNA material in the spermatophore. Each spermatophore was mounted on a microscope slide and stained with a DAPI solution (VECTASHIELD Mounting Medium with 4',6-diamidino-2-phenylindole 1,5 µg/ml), following (Dallai, Zizzari and Fanciulli, 2009). DAPI binds to the DNA and as the spermatophore coat (i.e. the external layer) of Collembola does not contain cell nuclei or DNA (Dallai, Zizzari and Fanciulli, 2009), blue coloration indicates the presence of sperm DNA material only. Pictures of the stained spermatophores were taken with a camera (ZEISS AxioCam ICc 3) mounted on a fluorescence microscope (ZEISS Axioskop 20) with a 1000x magnification, and the area of DAPI stained sperm droplet was measured (ImageJ, version 1.46; (Schneider, Rasband and Eliceiri, 2012).

Spermatophore attractiveness

To evaluate the attractiveness of spermatophores to females, we conducted two sets of two-choice trials using a Y-shaped glass olfactometer with vials attached to the proximal end of the choice arms (Fig. S2). First, we tested the ability of females of the sexual CR population to detect the presence of spermatophores. For this test, virgin receptive females of the sexual population (N=36) were offered a choice between a vial with spermatophores and an empty vial. Subsequently, we assessed whether sexual females were attracted to spermatophore of the asexual lines over empty vials (BE, N=39; EV, N=26). In the second series of two-choice trials, we tested weather females could differentiate between vials containing spermatophores of different strains: sexual CR vs. asexual BE (N=42), sexual CR vs. asexual EV (N=35), and asexual BE vs. asexual EV (N=23). Vials with fresh spermatophores were obtained by placing males in individual vials one day before

the start of the female choice trials. Before each trial, the vials containing spermatophores were mounted at the extremities of the choice arms of the olfactometers. Choices were scored when a female walked into one of the arms for more than 5 mm. Choice tests were terminated after 10 minutes, regardless of whether the female had made a choice. After the test, females were placed in a vial with spermatophores of a sexual male to check if they were receptive during the test.

Statistics

Lengths of reproductive and non-reproductive instars of the sexual CR population were compared using Wilcoxon rank sum tests with continuity correction. All three lineages of *F. candida* were tested for differences in stalk length, sperm droplet size and amount of sperm cell within the sperm droplet. Assumptions of normally distributed residuals and homogenous variance between the different treatments were tested with the Shapiro-Wilk normality test and Bartlett test of homogeneity of variances, respectively. Stalk length met both criteria and was therefore analysed with a one-way ANOVA. Sperm droplet size and amount of sperm did not meet the requirements of normality. Therefore, these traits were analysed using a non-parametrical Kruskal-Wallis rank sum test. Pairwise comparisons were obtained with a Wilcoxon rank sum test with Benjamini-Hochberg p-value adjustments. To test female preference in the behavioural assays we applied a Chi-squared test against a H_0 of no choice. Females that did not enter one of the arms within the observation time were excluded from statistical analysis. All tests were run in R (R Core Team, 2017).

Results

Reproductive biology of sexual F. candida

In order to characterize the reproductive traits of the sexual CR population of *F. candida* we monitored individual springtails daily for ten weeks (Figure 1). As expected, both males and females of the sexual population alternated reproductive with non-reproductive instars. Females took up a fresh spermatophore each reproductive instar and laid on average 31 (± 12) eggs in their first reproductive instar. The development time of the eggs was 10.1 (± 0.3) days. The sex ratio of the sexual clutches was 48.7 % males to 51.3 % females, which does not deviate significantly from an equal distribution (binomial-test, R, $P = 0.48$). Both the number of days as well as the number of instars until first reproduction differed between both sexes. Female age at maturity was 17.2 (± 1.6) days in the 6^{th} instar, whereas males developed faster and reached the first reproductive instar after 14.9 (± 1.4) days in the 5^{th} instar. The first male reproductive instar, with on average 172 (± 95)

spermatophores, was significantly longer than the subsequent non-reproductive instar (V = 465, $P < 0.000$). An increased number of spermatophores laid during the last days of the first reproductive instar was also recorded (Electronic Supplementary Information). In addition, female reproductive instars were longer when no spermatophores were available compared to when spermatophores were available (W = 470.5, $P < 0.000$).

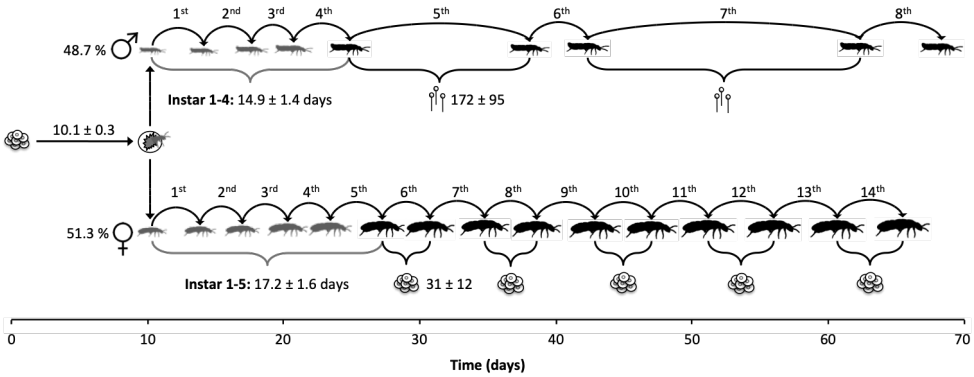


Figure 1: Reproductive biology of the sexual CR population of *F. candida* at 20 °C. Individuals keep growing through successive moults throughout their lifespan. For both males and females, the instars (intermoult periods) are numbered. After maturity, males and females alternate reproductive and non-reproductive instars. Juvenile instars are depicted in grey. Males produce spermatophores throughout the whole reproductive instar and females pick up a new spermatophore at the start of each reproductive instar to fertilise their eggs. Time intervals are given in days ± standard deviation (sd). Number of spermatophores for males and number of eggs for females during their first reproductive instar are given in number ± sd. The sex ratio is given in percentage.

Male sexual traits

Morphological characteristics of the spermatophore

To detect degradation patterns in male sexual traits, we measured stalk length, sperm droplet size and amount of sperm of male spermatophores of the three different populations. While the stalk length did not vary significantly among the three lineages ($F_{(2,39)} = 2.778$, $P = 0.0745$; Fig. 2A), there was a significant difference in sperm droplet size and amount of sperm (droplet size: $\chi^2(2) = 9.0464$, $P = 0.01085$; sperm: $\chi^2(2) = 29.371$, $P = 4.189e\text{-}07$; Fig. 2B-C). The asexual EV line had larger sperm droplets than the asexual line BE and the sexual population (CR). The amount of sperm with DNA differed significantly among all three lineages with both asexual lines showing significantly fewer fluorescent sperm heads than the sexual population, and the BE line showing less than the EV line (Fig. 2D).

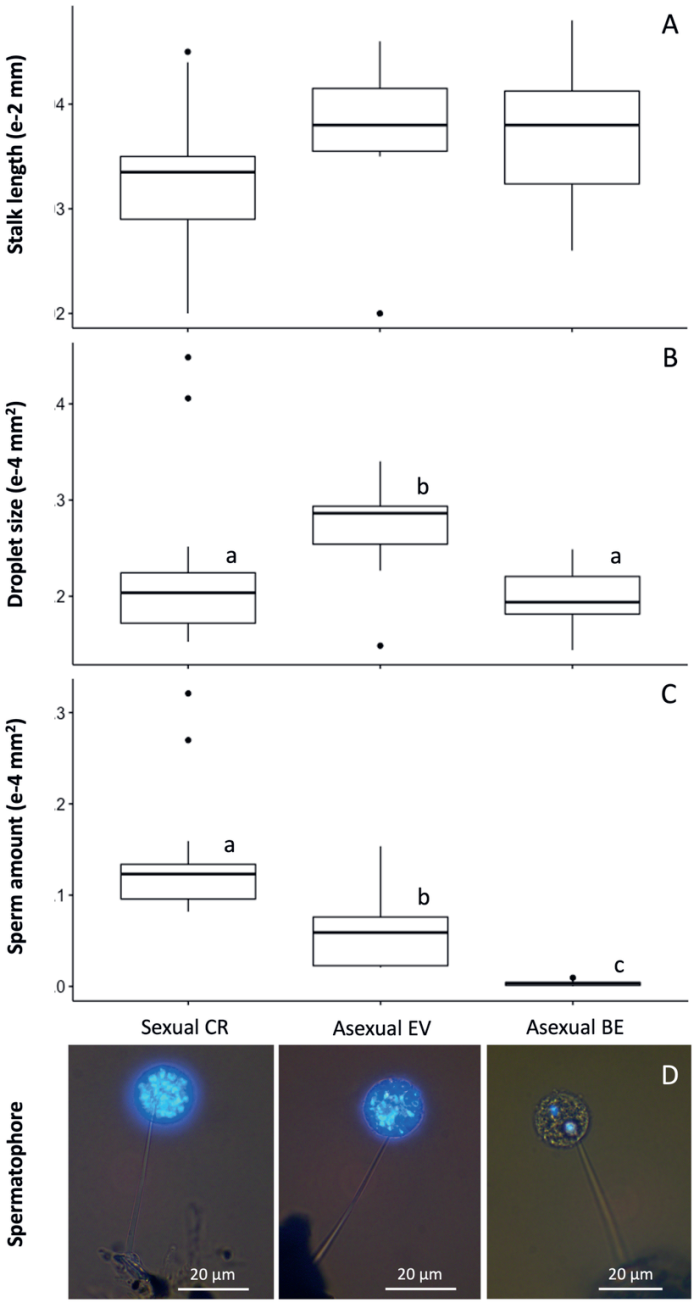


Figure 2: Spermatophore-related traits of three *F. candida* lines: stalk length (A), sperm droplet size (B), amount of sperm heads measured as DAPI stained area (C) and fluorescence microscope view of DAPI-stained spermatophores (D). Box plots depict the first to third quartiles with the median as inner line and whiskers showing the maximum and minimum, except for when outliers are present, in which case they show the inner fences.

Spermatophore attractiveness

In a series of two-choice tests we tested spermatophore attractiveness to receptive females of the sexual line CR. CR females were significantly more attracted to spermatophores of males of their own population than to empty vials ($\chi^2(1) = 16.03$, $P < 0.001$; Fig. 3). They also preferred spermatophores of the asexual BE line over empty vials ($\chi^2(1) = 8.76$, $P < 0.01$; Fig. 3) but did not show any preference for spermatophores of the asexual EV line ($\chi^2(1) = 1.09$, $P = 0.297$; Fig. 3). In a second two-choice trial, we tested whether sexual CR females differentiated between spermatophores of the different lines. Spermatophores of the sexual CR population were significantly more attractive to females than the asexual EV spermatophores ($\chi^2(1) = 4.5$, $P < 0.05$; Fig. 3), but CR females showed no significant preference for the CR spermatophores compared to the spermatophores of the asexual BE line ($\chi^2(1) = 0.64$, $P = 0.423$; Fig. 3). Also, no significant preference was detected between spermatophores of the two asexual lines ($\chi^2(1) = 0.47$, $P = 0.491$; Fig. 3).

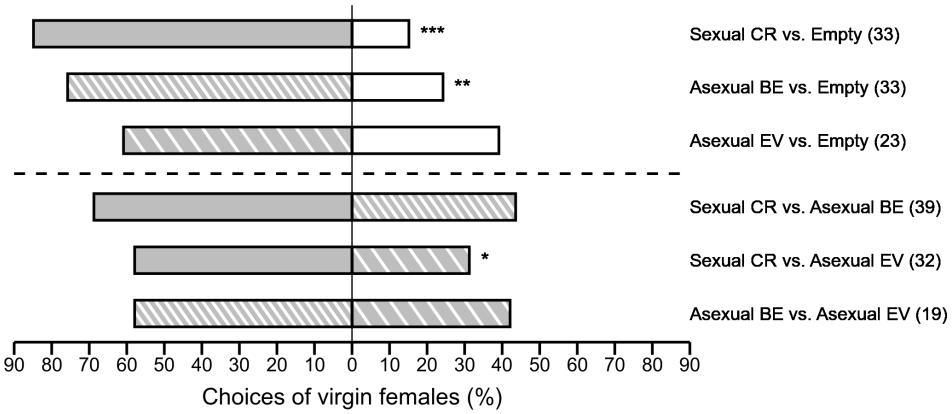


Figure 3: Behavioural responses of sexual *F. candida* females to spermatophores of different *F. candida* lines in a Y-olfactometer. Origin of spermatophores in the two-choice tests is shown on the right. Numbers in brackets indicate the absolute number of females that chose. Total number of tested females is given in the text. Asterisks indicate significant differences determined by a Chi-squared test for given probabilities (* is P<0.05, ** is P<0.01, *** is P<0.001). Inside bar patterns indicate the average amount of sperm for each lineage: sexual CR > asexual EV > asexual BE.

Discussion

In this study, we investigated male sexual trait decay in the asexual springtail *F. candida* by comparing spermatophore-related traits of two asexual lines and one sexual population, whose reproductive biology is described here for the first time. In the asexual lines, we found evidence for decay of the amount of sperm within the sperm droplet and of spermatophore attractiveness. Moreover, our results suggest that the extent of decay of such male sexual traits differs between the two asexual lineages.

Reproductive biology of sexual F. candida

While the reproductive traits of asexual *F. candida* are well-documented (Snider, 1973; Fountain and Hopkin, 2005; Krogh *et al.*, 2008), those of sexual populations have not received the same attention. In this respect, our study provides useful information on the sperm transfer mode and on the reproductive behaviour of *F. candida*.

Sexual males deposit spermatophores in the absence of females, thus confirming a dissociated sperm transfer system. Males reach maturity on average two days before females and the reproductive instar is longer than the non-reproductive instar. Faster maturation of males is common in Collembola and other invertebrates (Poggendorf, 1956; Waldorf, 1971, 1975; Zizzari, van Straalen and Ellers, 2016). A longer male reproductive instar with an increase in spermatophore deposition toward its end (Fig. S1), as found in *F. candida*, has also been observed in other Collembola (Testerink, 1982). Prolonging the reproductive instars could be a way to maximize male reproductive success, as it gives a male a longer time window in which spermatophore deposition can occur (Waldorf, 1971). Females appear to have plastic reproductive instar lengths, as we found the reproductive instar to be prolonged in the absence of spermatophores. In two-choice trials, receptive females showed a strong preference for vials containing spermatophores over empty vials. These findings confirm that also females of *F. candida* locate spermatophores based on chemical cues and add to the increasing evidence that collembolan spermatophore-associated pheromones mediate partner attraction (Zizzari *et al.*, 2017). Our study shows that egg development time, age at first reproduction, and number of juvenile instars were within the range observed in females of asexual lines (Snider, 1973; Fountain and Hopkin, 2005; Krogh *et al.*, 2008). However, the average clutch size of sexual females was lower than observed in asexual females (Fountain and Hopkin, 2005; Krogh *et al.*, 2008).

In conclusion, a continuous moulting throughout the life span, the alternation of reproductive and non-reproductive instars, and female attraction to spermatophore odour, confirm that the reproductive biology of sexual *F. candida* is in line with what is observed in other collembolans species. Overall, this life-

history information makes an important laboratory model available for further laboratory investigation on the transition from sexual to asexual reproduction promoted by *Wolbachia* and on the evolutionary dynamics of sexual traits loss.

Decay of male traits in asexual F. candida

Spermatophore-related traits

We found a strong indication that in asexual *F. candida*, sperm production is subject to decay, as in both asexual lines the amount of sperm DNA material was lower compared to the sexual population. In the spermatophores of one of the asexual lines even, fluorescent sperm heads were almost completely absent. Our results are consistent with previous findings as sperm production has been indicated to be one of the first male traits to deteriorate in asexual lines and such a trend is reported in several species (van der Kooi and Schwander, 2014).

In addition to compromised sperm production another male sexual trait was observed to be degraded: spermatophore scent. We evaluated the olfactory attractiveness of spermatophores to sexual females in two-choice trials. Spermatophores of the sexual line were significantly more attractive than the asexual EV spermatophores, while females showed no preference for sexual spermatophores over asexual BE spermatophores. The observed reduced attractiveness for asexual EV spermatophores may be due to decay of the pheromone biosynthetic pathway, leading to a reduced or modified sex pheromone expression. A variation of the pheromone profile, owing to the absence of directional selection, could also explain the reduced spermatophore preference, as fine-tuning of sexual signals and reception can only be maintained under strong selection. Given that these signals experience weak selection in asexuals, alteration of the expression of formerly adaptive olfactory cues is expected. Alternatively, population-specific mating preferences, as shown for instance in the parasitoid wasp *Leptopilina clavipes (Kraaijeveld et al., 2009),* could be hypothesized. Thus, a difference in spermatophore attractiveness may have been due to differences in sex pheromone profile among populations. The characterization of the pheromone bouquets emitted by the spermatophores of the three investigated populations would help elucidating the observed patterns in spermatophore preference.

A shift in the sperm droplet size was detected in only one of the asexual lines, with an increase in size rather than the typical decrease associated with trait decay. Thus, the sperm droplets of males of the asexual line EV were larger than those of the other two examined lineages. However, the size of sperm and spermatophores often co-evolves with width and size of the female reproductive tract (Presgraves, Baker and Wilkinson, 1999; Dallai *et al.*, 2014). For this reason, a divergence in the size of a spermatophore, either larger or smaller, is likely an

indication of weak co-evolutionary constraints that determine the optimal size of the spermatophore to be picked up by a female.

We tentatively suggest that male sexual traits have not decayed in parallel in the two asexual groups. Although, the difference in spermatophore attractiveness could be a consequence of lineage diversification, the variation in the amount of sperm in the sperm droplets is unlikely to stem from divergent selection between populations. Rather, it could stem from sexual trait decay in the males.

If the differences in spermatophore attractiveness, sperm droplet size, and amount of sperm are indeed a consequence of decay, the question arises as to which evolutionary forces promote a different rate of decay. Mutation accumulation is the result of a stochastic process (Kimura, 1983); hence, progression of trait decay is expected to differ between asexual lines. Conversely, if a sexual trait is subject to strong regression in old asexual lines as well as in young ones, as was shown in asexual female *Timema* (Schwander et al. 2013), the causative evolutionary process is very likely negative selection. Although we do not know the number of independent transitions of the asexual lines, the pattern of trait decay found in our study appears to be consistent with neutral mutation accumulation rather than selection.

Since trait decay by neutral mutation accumulation can be found in long-lived asexual lineages only (Kimura, 1983; Hall and Colegrave, 2008; van der Kooi and Schwander, 2014), we suggest *Folsomia* to be an ancient asexual. Male trait decay was detected in oribatid mites which are estimated to having been parthenogenetic for 100 million years (Heethoff *et al.*, 2007; van der Kooi and Schwander, 2014). The high genetic differentiation between sexuals and two Italian asexual lines (Frati *et al.*, 2004), might support ancient asexuality in *Folsomia,* although genetic variation can also be associated with ecological niche differentiation, as shown in asexual grass thrips (van der Kooi et al 2019).

Future studies should shed light upon the intraspecific molecular divergence between the *Folsomia* sexual and asexual lineages BE and EV to estimate for how long they have been parthenogenetic. Moreover, further investigation should ascertain whether spermatophores of asexually produced males contain viable sperm and whether *Wolbachia*-infected females can produce offspring with spermatophores of sexual males.

Acknowledgements

We thank Janine Mariën and Rudo A. Verweij for maintaining the cultures of asexual *F. candida* lines and Prof. P. Paolo Fanciulli for providing us with the sexual population of *F.candida*. Finally, we are grateful to the anonymous reviewers for insightful comments.

Data Availability

All data generated or analysed during the current study are included in this published article and its Supplementary Material.

Compliance with Ethical Standards

Conflict of interest: the authors declare that they have no conflicts of interest.

Supplementary Material

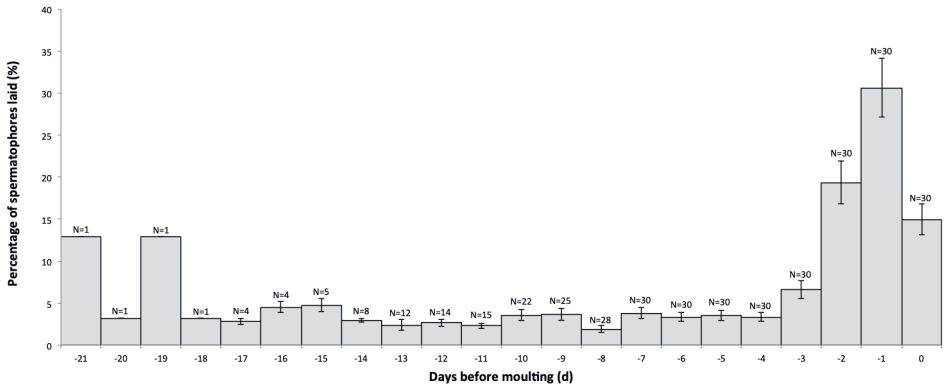


Figure S1: Male spermatophore production in F. candida sexual line during the first reproductive instar. The figure shows the percentage of spermatophores laid (± SE), for each day prior to moult. The numbers of males that could be included per day is shown above each bar.

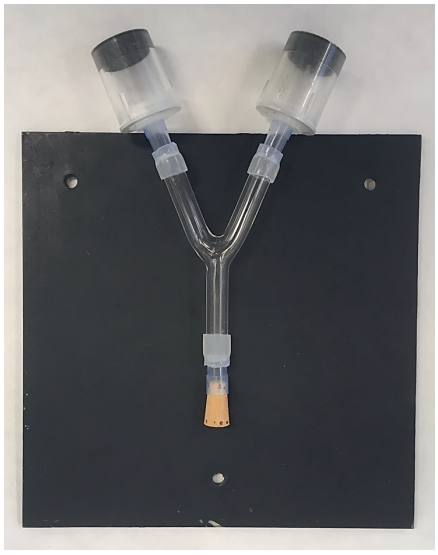


Figure S2: The response of *F. candida* female to different odour source was tested in a glass Y-tube olfactometer (1 cm inner diameter) with an introduction arm (6 cm) and two choice arms (6 cm). The angle between the two choice arms was 50. Individual females were placed into the introduction chamber from which they locomoted to the Y-intersection and then selected a Y-arm to enter. Two scent sources (i.e., Plexiglas vials) were attached to the proximal end of the choice arms. A fine mesh dived the vials from the olfactometer. The set-up could be disassembled for cleaning between trials.

CHAPTER 3



DKJR

Transinfections of *Wolbachia* between *F. candida* lines; heat stress, re-infection and potential effects of bacteriophage WO

A.A. Kampfraath*, J.M. Brouwer*, J. Mariën, K. Kraaijeveld and J. Ellers

Status: unpublished *shared first author

Abstract: *Wolbachia* is a genus of endosymbiotic bacteria that infects invertebrates and causes reproductive alterations. In the springtail *Folsomia candida Wolbachia* (*w*FoI) has been linked to parthenogenesis, but its exact role remains unclear since elimination of *w*FoI leads to sterility and transinfection experiments with *w*FoI on sexual *F. candida* did not generate stably infected lines. Many Wolbachia, like *w*FoI, also contain the temperate bacteriophage WO, which has been shown to be activated by stress resulting in lysis of *Wolbachia* in other hosts.

Here we investigate the effect of temperature and transinfection on the lytic activity of the phage. We measured the phage/*Wolbachia* ratio by Q-PCR during heat stress of *Wolbachia*-infected *F. candida* at 30°C for 19 days, and after transinfecting *Wolbachia* into naturally uninfected sexual *F. candida* and into parthenogenetic *F. candida*. Additionally, we measured egg laying and hatching rate of springtails cured of *Wolbachia* with rifampicin and subsequently re-infected with *w*FoI.

Our results show a significant decrease of *Wolbachia* densities over time in the heat treatment, but no significant difference was found in phage/*Wolbachia* ratios between the heat-treated springtails and controls. Transinfection into the sexual strain was unsuccessful with almost no *Wolbachia* or phage DNA found, while transinfection into the parthenogenetic strain caused a significant increase in *Wolbachia* densities compared to the control, while the phage/*Wolbachia* ratios remained unchanged. Therefore, unsuccessful transinfection into the sexual strain was most likely caused by an immune response of the sexual strain. Finally, the number of eggs per clutch of the springtails under antibiotic treatment was significantly lower than the controls and the hatching rate was even reduced to zero. After discontinuing the antibiotics treatment or re-infecting the springtails with *Wolbachia*, few eggs did succeed in hatching. These results support the hypothesis that *Wolbachia* is important for the parthenogenetic reproduction of *F. candida*, but do not elucidate the role of the phage WO in the infection process.

Keywords: Bacteriophage WO, *Wolbachia*, heat-stress, transinfection, parthenogenesis

Introduction

Symbiosis, defined as "the living together of unlike organisms" (De Bary, 1879) has played an important role in major evolutionary events such as the origin of the eukaryotic cell (Archibald, 2015; Keeling and McCutcheon, 2017). Species living in symbiosis have close interactions and some of the most intertwined interactions are found amongst endosymbionts that live inside their hosts' cells. *Wolbachia* is a common bacterial endosymbiont, that is found in a wide range of invertebrate and nematode species and is estimated to potentially infect 66% of all arthropod species (Hilgenboecker *et al.*, 2008). The nature of its interaction with its host can differ strongly, as in nematodes they are mutualistic and, in some cases, provide means to function fully, while in arthropods they interfere with the reproductive system (Werren, Baldo and Clark, 2008; Kent and Bordenstein, 2010). These reproductive manipulations include: feminization, where genetic males develop into females; male killing, where males are feminized and die during larval development; cytoplasmic incompatibility, in which sperm from *Wolbachia*-infected males is incompatible with eggs from females not carrying the same strain; and induction of parthenogenesis (PI), causing unfertilized eggs to develop into females (Werren, Baldo and Clark, 2008).

While PI by *Wolbachia* has mostly been found in haplo-diploid species, it has also been found in the diploid-diploid springtail *Folsomia candida*. Its parthenogenetic lineages are infected with *Wolbachia,* while sexual lineages are not (Frati *et al.*, 2004; Kampfraath *et al.*, 2020). However, unlike haplodiploid species where the link between *Wolbachia* infection and PI is unequivocal (Russell and Stouthamer, 2011), in *F. candida* the exact role of *w*FoI (the *Wolbachia* lineage infecting *F. candida*) in PI is not yet fully understood. *w*FoI has been observed to aggregate around cytoplasmic asters in the eggs of *F. candida* during meiosis (Riparbelli, Giordano and Callaini, 2006) and eggs of females cured from *w*FoI by heat or antibiotics stop developing, leaving them unhatched (Pike and Kingcombe, 2009; Timmermans and Ellers, 2009). Still, it cannot be completely ruled out that the lack of egg hatching is caused by the stress of the curing treatments rather than the absence of *Wolbachia.* Due to the infertility of cured females, no cured lines can be established to falsify this.

To get a better understanding of the interaction between *w*FoI and its host, we attempted to create a newly infected line of *F. candida* through the transfer of *w*FoI into the females of the sexual line. Artificial horizontal transfer of *Wolbachia* by transinfection has been used to control arthropod-borne diseases and investigate *Wolbachia*-host interactions (Hughes and Rasgon, 2014). We attempted to establish a new infecting in several ways. In our first trial, a mixture of crushed springtails in SPG buffer was transferred (n=173), which led to high losses directly after the transfer procedure. In a second trial we directly transferred haemolymph from the sexual line into the parthenogenetic line (n=200). Within this trial the

mortality after the transfer procedure was more moderate and 16 females managed to produce eggs. We measured the Wolbachia densities in these injected females and their clutches by qPCR. Only 4 females had high Wolbachia titres and in only 10 of the 34 clutches signs of *Wolbachia* titres were visible within the 5 selected F1 females. No eggs were laid by the remaining F1 females, thus our efforts to establish a new parthenogenetic line failed and further tests on the interaction between *w*Fol and its host could not be done. This demonstrated that establishing a new infection is not trivial and suggested that the sexual line might be less susceptible to wFol or that the conditions for establishment were not met.

Bacteriophages might also have played a role in the interaction between *F. candida* and *w*Fol, making this a tripartite symbiosis (Kampfraath *et al.*, 2019; Bordenstein and Bordenstein, 2022). These bacteriophages have been found in many other *Wolbachia* (Bordenstein and Bordenstein, 2022) and within wasps there is evidence that phage and bacterial densities are inversely related, suggesting that they go into the lytic cycle and thereby kill *Wolbachia* cells (Bordenstein *et al.*, 2006). Experimental work has demonstrated that their lytic cycle can be triggered when the host of the phage encounters stress, such as temperature and host age (Bordenstein and Bordenstein, 2011; Lu, Zhang and Hong, 2012). Moreover, a change in relative phage density was seen when Wolbachia was transferred between species, although in this case the phage density decreased (Chafee *et al.*, 2011). Possible explanations, such as lower stress for *Wolbachia*, stochastic changes, or tritrophic interactions between the phage, *Wolbachia* and its new host, do not yield unequivocal predictions of these phage dynamics. Thus, the reason for this relative decrease in phage numbers after a host change is still unclear.

In this study we aim to determine whether the lack of establishment of *Wolbachia* within the sexual line of *F. candida* could be linked to activation of lytic cycle of the phages within its bacterial genome. To address this question, we measured the amount of *Wolbachia* and phage WO after transfer from the parthenogenic line into the sexual line. Activation of phages by transinfection has not been reported before, therefore we also examined whether heat stress treatment can cause the phages within *w*Fol to go into the lytic cycle and affect the relative abundances of Wolbachia and phage WO. In addition, we tested if curing and reinfection of the parthenogenetic line influences egg hatching to gain stronger evidence of *Wolbachia*'s role in the parthenogenetic reproduction of *F. candida.*

Material and methods

Folsomia candida cultures

A sexual and a parthenogenetic line of *F. candida* from the culture of Vrije Universiteit Amsterdam were used. The sexual lineage was originally collected in

Cremona, Italy (Frati *et al.*, 2004; Kampfraath *et al.*, 2020) and the parthenogenetic springtails were taken from the VU1 and VU2 cultures, which are derived from the well-characterized asexual "Berlin" strain that is commonly used for lab tests. The parthenogenetic culture was kept at 16°C, 75% humidity, and 16:8-hour light-dark cycle, and was fed dry baker's yeast (Algist Bruggeman N.V., Belgium). The sexual culture was kept in a separate climate room at 20°C, 75% humidity, and 16:8-hour light-dark cycle, to avoid mixing of the cultures. They were also fed dry baker's yeast (Dr. Oetker, Netherlands).

Synchronizations

To create *F. candida* cohorts of the same age three synchronisations were done. To start a synchronisation, adult females were placed in plastic jars filled with a 1 cm layer of moist black plaster of Paris (1 part charcoal and 9 parts plaster of Paris). After 48 hours the adults were removed and the remaining eggs were incubated at 20°C, 75% humidity, and 16:8-hour light-dark cycle. After hatching at a similar time, creating a cohort of *F. candida* of the same age, the juveniles were fed with dry baker's yeast (Dr. Oetker) twice a week and plaster was kept moist. For the heat treatment experiment, a synchronization with 8 jars containing 25 adult female parthenogenetic *F. candida* was done. A second synchronization for the transinfection, curing and re-infection experiments consisted of 20 jars containing 25 parthenogenetic *F. candida* females. Finally, a third synchronization was prepared with the sexual line of *F. candida* and consisted of ten jars with 25 females and 17 (±2) males each. Sexual and parthenogenetic synchronizations were kept in separate climate rooms.

Heat treatment

Before starting the actual heat treatment experiment, a preliminary test was done with 17-day-old synchronized springtails. To avoid drying out of the plaster during the heat treatments, containers with sloped plaster were created that could be continuously moisturized by a surplus of demineralized water at the lower side, while still keeping an unflooded surface for the springtails to live on. The springtails were put in the jars in a water bath at 34°C for four hours per day for three days. One control animal was frozen in the freezer at -20°C before the test and five treated animals were frozen in the freezer after one day and three days. This test was used as a preliminary dataset to get some insight in the effect of the higher temperature at the beginning of the experiment.

For the actual heat treatment experiment, eight plastic containers (250 ml) with sloping plaster were prepared (see Figure 1A). Juveniles of 17 days old from the synchronization were then 1 to 1 transferred to these containers, put in a climate chamber at constant 30°C, 95% humidity and 16:8-hour light-dark cycle. A

control of around 20 springtails per synchronization jar were put together and kept in the climate chamber at 20°C. All jars were watered every two days, and dry baker's yeast was provided once a week. Another preliminary test showed that only after day eight a strong decrease in *Wolbachia* density occurred. Therefore, 15 heat-treated springtails were sampled on day 8, 11, and 15, while an additional 8 individuals were sampled on day 19. On the first timepoint they were frozen in a freezer at -20°C, on the later timepoints samples were snap-frozen in liquid nitrogen, as this would reduce the risk of degradation of bacteriophages. 15 control springtails were sampled on day 8 and 11 to cover both freezing methods.

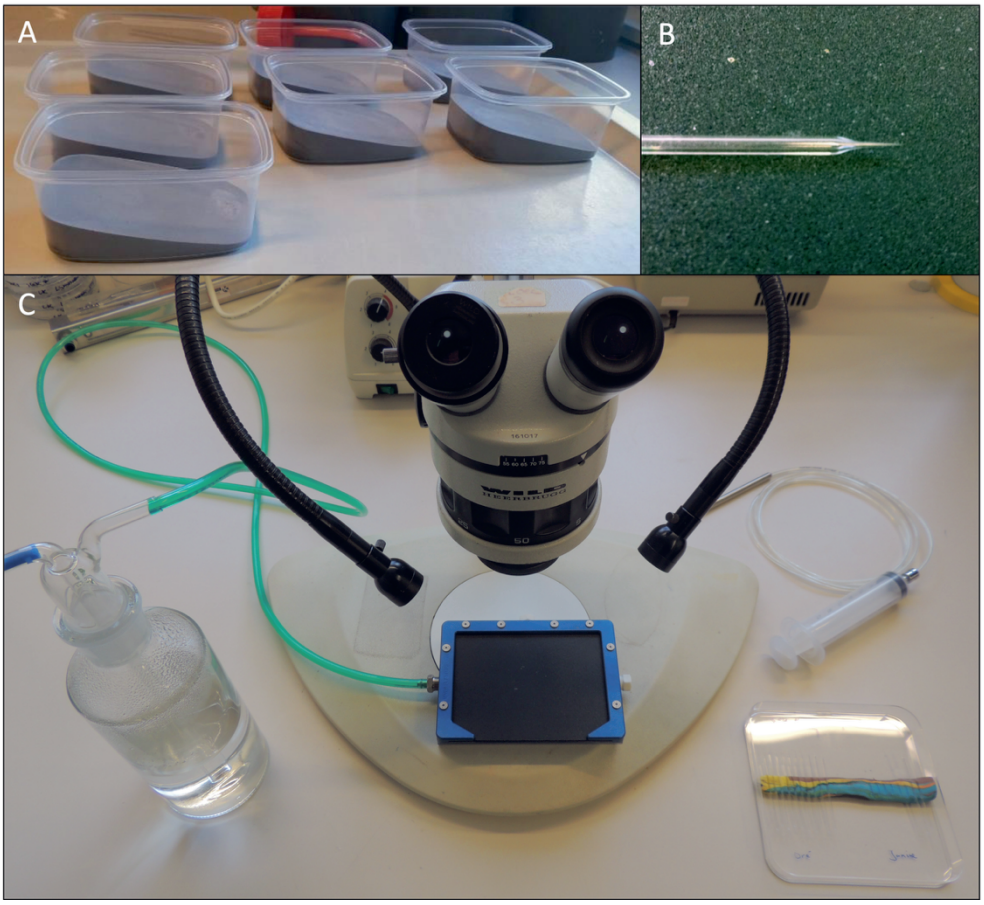


Figure 1: Pictures of the experimental set-up: A) Containers with sloped plaster of Paris B) Close-up of pulled glass capillary needle C) full transinfection set-up with a bubbler, flypad, binocular microscope, a syringe with needle holder and the pulled glass capillary needles.

Wolbachia transinfections

Synchronized 17 to 21 days old females of the parthenogenetic and sexual lines were injected with fresh haemolymph acquired from asexual females on consecutive days. Fresh haemolymph was taken from the ventral side of the abdomen of a large parthenogenetic female from the main culture and was then injected in the abdomen of the recipient. For each transinfection, a new donor and new hollow glass needle were used, connected to a 30 ml syringe by a plastic tube to control the pressure (see Figure 1B and C). Fifteen springtails were randomly sampled and frozen in liquid nitrogen at four time points (day 0, 1, 3 and 7). For the parthenogenetic line only the first three timepoints were sampled. As an additional control 15 untreated 17-day-old female springtails of both the sexual and parthenogenetic line were frozen in liquid nitrogen.

Quantitative PCR analysis and DNA extraction

To measure the quantities of *Wolbachia* and phage WO DNA in the samples, two Q-PCRs were developed. *Wolbachia*-quantification was based on PCR-amplification of a 101bp region of the single copy gene gltA of *Wolbachia* with the following primer pair: *glt*Af CTCTCCATGCTGATCATGAACA and *glt*Ar TAACCACTCCTGCAGAAAGGC. To detect the WO phages, a 120bp intergenic region that was present eight times in the regions of phage origin of the *w*Fol genome was selected. This region covered all three bacteriophage WOFol regions and largest WO island (two times in WOFol1, two times in WOFol2, three times in WOFol3 and once in WOI1, Figure 2) and was amplified with the primer pair (*WOFol*f CAAACTCCTATACGCGGTCACC, *WOFol*r CTTGAGGCTCCCTCTGCTTTTAT).

To get correct estimates of copy numbers with qPCR analysis, two calibration curves were created. DNA extracted from one adult parthenogenetic springtail from the culture was diluted five times in a ratio of 1 part sample and 3 parts sigma water, with relative dilution values of 1024 for the non-diluted sample and 256, 64, 16, 4, and 1 respectively for the dilutions. These dilutions were then subjected to quantitative PCR analysis, using the CFX-96 real time thermal cycler. The dilutions were measured in duplicate for both *Wolbachia* and the bacteriophages using 4 µl template DNA, 0.5 µl of the forward and reverse primers, 10 µl BioLine Sybr Green, and 5 µl of sigma water in a total volume of 20 µl. These calibration curves were used to calculate a ratio of relative number of bacteriophages to number of *Wolbachia* cells of the heat-treated and injected samples. This ratio was used as a measurement for the activity of lytic bacteriophages within the samples of the heat-treated and injected springtails. The samples of each treatment were spread out over multiple wells-plates with at least one control sample per plate. Each wells-plate was split in three compartments: in the first compartment the samples were measured with the bacteriophage primers,

in the second compartment the same samples were measured using the *Wolbachia* gltA primers, and in the third compartment a calibration curve was added of known concentrations of *Wolbachia* gltA copy numbers, integrated in pGEM-T plasmids, used to calculate the absolute starting quantity (ASQ) of *Wolbachia* in the samples.

All DNA was extracted using the Promega SV DNA isolation method for springtails. The springtail was rinsed with 70% ethanol and then homogenized in 100 µl PBS. Then, after adding 100 µl NLS buffer and 2 µl proteinase K (10 mg/ml), the sample was incubated at 55°C overnight. After incubating, 170 µl DNA lysis buffer was added and the sample was centrifuged. The supernatant was then washed three times with 500 µl wash buffer, incubated for one minute at room temperature, and the DNA was dissolved in 50 µl of sigma water. DNA samples were then subjected to quantitative PCR, using the CFX-96 and the same method as described in the previous paragraph.

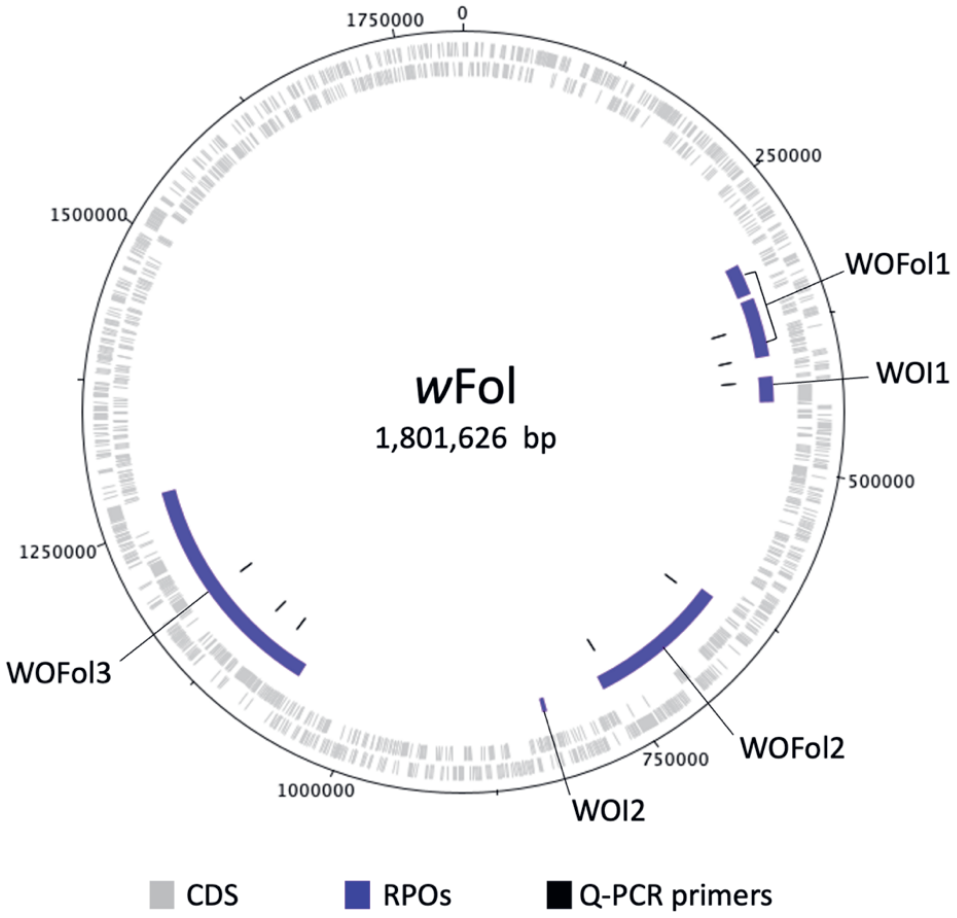


Figure 2: Locations of WOFol Q-PRC primers within the *w*Fol genome. Regions of phage origin (RPO) are in purple, coding sequences in grey and Q-PCR primers in black.

Curing and re-infection treatment

After hatching of the eggs in the synchronisation, 15 jars were assigned to the curing treatment, in which the juveniles were provided with a mixture of antibiotic rifampicin, dry baker's yeast and demineralized water (1 mg rifampicin: 100 mg yeast: 200 µl demineralized water) on a small disk of filter paper (Ø5mm) and the other 5 jars were used as control and were fed with a mixture of dry baker's yeast with demineralized water (100 mg yeast:200 µl demineralized water) on filter paper, both prepared weekly and stored at 5°C .

For the experiment, 30 control springtails and 80 treated springtails of 28 days old were set apart in small cylinders (4 cm high, Ø1.5 cm) with a 0.5 cm layer of moist black plaster of Paris. These were placed in large plastic containers (1000 ml) filled with a 1 cm layer of moist white plaster of Paris, to keep the springtails hydrated. Control springtails were fed with the non-spiked food mixture twice a week until they laid their first egg clutch. After that, the adults were transferred to a new cylinder, fed again with non-spiked food, and allowed to lay a second egg clutch. The treated springtails were divided into two groups: 30 springtails to be cured, and 50 to be re-infected with *Wolbachia*. The cured springtails were fed with antibiotics-spiked yeast until the first egg clutch, after which they were transferred to a new cylinder, and fed with non-spiked yeast. Of the 30 cured springtails, one was excluded as it had gotten the wrong food. The re-infection group was also fed with the antibiotics-spiked yeast until their first egg clutch and then transferred to a new cylinder. One turned out to be a parthenogenetic male and was therefore unable to lay eggs and removed from the experiment. After being fed with non-spiked food for two days to get rid of the antibiotics in their system, they were injected with *Wolbachia* using the same method as in the transinfection experiment. After the injection, they were again fed with the non-spiked food until their second egg clutch.

All eggs were incubated at 20°C. When eggs were overgrown by mould or did not hatch after 14 days, they were counted by hand under a microscope, after which the egg clutch was removed. When the eggs did hatch, the number of unhatched eggs was counted under the microscope, and the juveniles were fed dry baker's yeast until they were between 7 and 14 days old. Then the jars with the juveniles were photographed, and the number of juveniles in the photo was counted using ImageJ.

Data analysis and statistics

Relative numbers of bacteriophages to *Wolbachia* cells in the samples were calculated based on calibration curves of WOFol and gltA respectively, which were expressed as exponential trend lines with the following equations: *Wolbachia* (gltA)

ASQ = 1E +09$e^{-0.661Cq}$ and phage RSQ = 9E+07$e^{-0.655Cq}$, where ASQ is the absolute starting quantity of *Wolbachia* and RSQ is the relative starting quantity of phages in the sample and *Cq* the measured cycle where the threshold line of 1000 is crossed. The calculated RSQs of the phage and *Wolbachia* gltA were then used to calculate the ratio of relative number of bacteriophages to absolute number of *Wolbachia* cells. The ASQ of *Wolbachia* cells in the samples was added on each well plate and values were then transformed using a $\log_{10}$ transformation for the statistical part of the analysis, because with the untransformed data the assumptions of normally distributed residuals and homogeneity of variance of the model were violated.

The statistical analysis used for the bacteriophage/*Wolbachia* ratio and $\log_{10}$ of *Wolbachia* gltA copy numbers of the heat treatment was a One-Way ANOVA using the lme function (package nlme) in R version 3.2.5. Wells-plate was added as a random factor, to control for differences in the prepared reaction mixture between plates, and "Treatment" (the two controls and the different days of the heat treatment) was added as factor. Furthermore, a Tukey post hoc test was done to test which days differed from each other, using the function glht (package multcomp).

The same analysis was used for the bacteriophage/*Wolbachia* ratio and $\log_{10}$ of *Wolbachia* gltA copy numbers of the parthenogenetic injected *F. candida*, with "treatment" (frozen immediately, frozen after one day, frozen after three days) as factor, and wells-plate as random factor to again control for differences between plates. Then, a Tukey post hoc test was performed to test differences between treatments, using the glht function. The sexual injected *F. candida* could not be tested, as quantitative PCR could neither detect bacteriophages nor *Wolbachia* gltA within these samples.

The statistical analysis for the number of eggs that were laid and the percentage of juveniles that hatched from the eggs was done with linear mixed-effects models, using the function lme (package nlme) in R version 3.2.5, with "treatment" (control, cured, re-infection) as factor and "individual" as repeated measure across clutch. A Tukey post hoc test was performed using the function glht to compare the treatments with each other.

As the preliminary heat test had a low number of replicates (N=5 for both days) and the temperature and duration differed from the larger heat experiment, no statistical analysis could be performed either within this test or between the two experiments.

Results

Influence of stress: heat treatment

The effect of heat on the bacteriophages and *Wolbachia* densities is presented in Figures 3 and 4. A significant decrease in *Wolbachia* gltA copy numbers was found in the heat-treated *F. candida* of the 19-day heat treatment, $F(5; 73) = 42, 60$; $p < 0.001$ (Figure 3B). As shown by the Tukey post hoc test, the heat-treated springtails on day 8 had significantly lower gltA copy numbers than the control of day 8, $z = -7.73$; $p < 0.001$. Also, the treated springtails of day 11, day 15 and day 19 had significantly lower *Wolbachia* copy numbers than the control of day 11, $z = -5.91$; $p < 0.001$; $z = -8.90$ and $p < 0.001$; and $z = -9.01$; $p < 0.001$ respectively. Furthermore, a decreasing trend in *Wolbachia* copy numbers is visible across the days of the 19-day heat treatment (Figure 3B): a significant decrease in gltA copy numbers was found between the treated springtails of day 8 and 15, and day 8 and 19, $z = -3.81$; $p < 0.01$ and $z = -4.75$; $p < 0.001$ respectively. Also, a significant decrease in gltA copy numbers was found between the treated springtails of day 11 and 15, and day 11 and 19, $z = -3.00$; $p = 0.03$ and $z = -4.08$; $p < 0.001$ respectively. However, there was no significant difference in *Wolbachia* gltA copy numbers between the treated springtails of day 15 and 19, $z = -1.58$; $p = 0.61$. In addition, no significant difference was found between the phage/*Wolbachia* ratios of the different treatment groups during the 19-day heat treatment ($F(5, 73) = 1.58$, $p = 0.18$) (Figure 4B).

Influence of stress: transinfection into a new host

The data of the transinfection experiment are shown in Figures 5 and 6. A significant difference between the treatments was found in *Wolbachia* gltA copy numbers of the injected parthenogenetic springtails ($F(3, 51) = 14.53$, $p < 0.001$) (Figure 5). The Tukey post hoc test indicated that day 0, day 1 and day 3 all had significantly higher values than the control, $z = 4.22$; $p < 0.001$ and $z = 5.81$; $p < 0.001$ and $z = 6.16$; $p <$ 0,001 respectively. No significant differences were found in phage/*Wolbachia* ratios of the parthenogenetic injected *F. candida* between the treatments, $F(3; 51) = 2.07$; $p = 0.12$ (Figure 6).

In the injected sexual *F. candida*, almost no *Wolbachia* gltA could be detected. In only two sexual samples, DNA was amplified in both replicates, with amplification starting in cycle 35.23 (±0.44). Also, in only five sexual samples, phage DNA was detected in both replicates, with amplification starting on average in cycle 34.95 (±1.17). However, the 10^1 *Wolbachia* gltA copy number samples of the calibration curve were amplified between cycle 30 and 33 and the 10^0 sample could not be detected. Thus, this could suggest either very low numbers of *Wolbachia* (>10) or that this is random amplification of other genetic material. Therefore, it could also be considered below the threshold of the detection level.

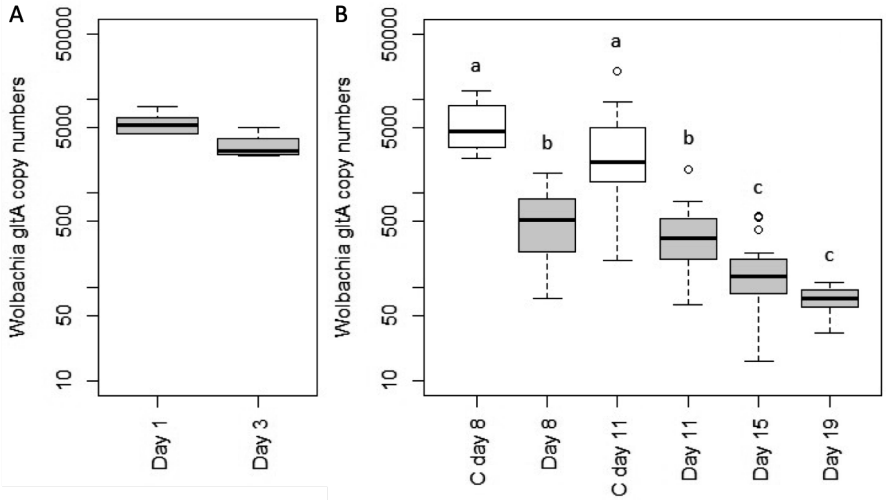


Figure 3: *Wolbachia* gltA copy numbers of parthenogenetic *F. candida* of the preliminary test (3A; N=5 per treatment) and 19-day heat treatment (3B; N=15 per treatment). Springtails in the preliminary test (Day 1 N=5, Day 3 N=5) and control and heat -treated till day 8 were frozen in the freezer, the remaining in liquid nitrogen. The boxplots show the median and first and third quartile, with whiskers of 1.5 IQR and the outliers are indicated by a circle. White boxes are controls and treatments and in grey. The letters indicate significant difference: $F(5; 73) = 42, 60$; $p < 0.001$.

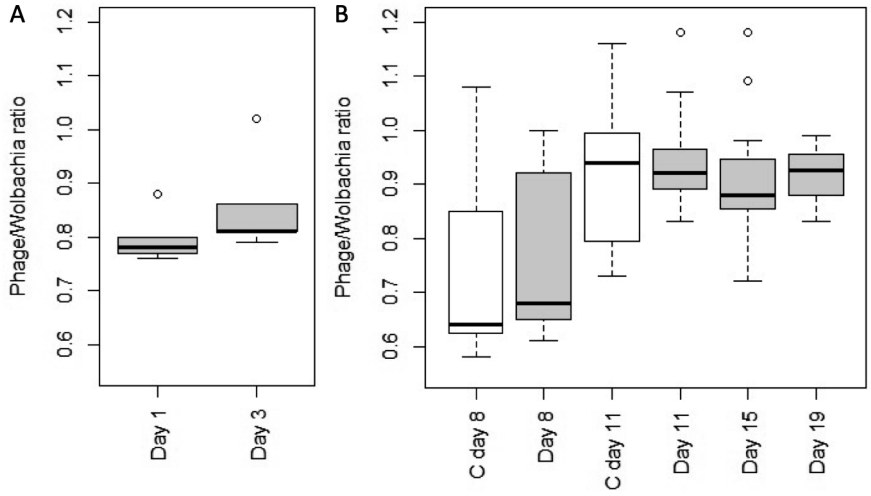


Figure 4: Phage/*Wolbachia* ratios of parthenogenetic *F. candida* of the preliminary test (2A; N=5 per treatment) and 19-day heat treatment (2B; N=15 per treatment). Springtails of the preliminary test and control and heat-treated till day 8 were frozen in the fridge, the remaining in liquid nitrogen. The boxplots show the median and first and third quartile, with whiskers of 1.5 IQR. White boxes are controls and treatments and in grey. Outliers are indicated by circles. 19-day heat treatment: $F(5, 73) = 1.58$; $p = 0.18$. Preliminary test: no statistical test performed.

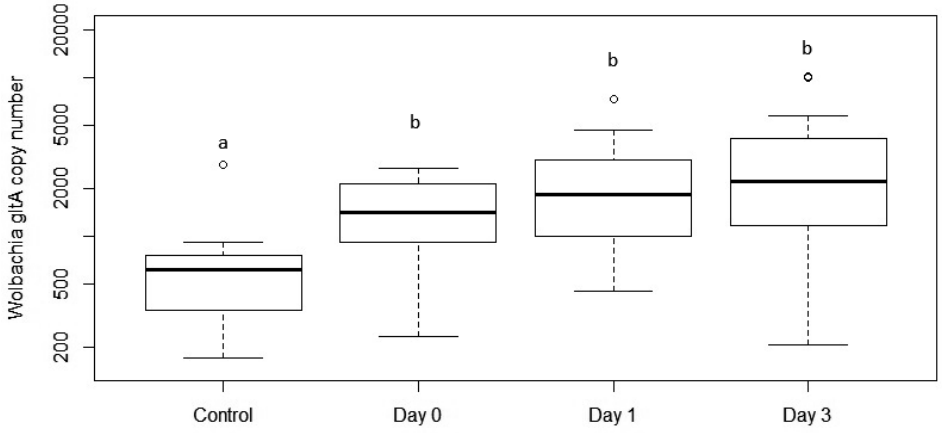


Figure 5: *Wolbachia* gltA copy numbers of parthenogenetic *Folsomia candida* injected with *Wolbachia*- containing haemolymph. The boxplots show the median and first and third quartile, with whiskers of 1.5 IQR. Circles indicate outliers and letters indicate significant differences. Measurements were taken immediately after the injection (day 0), after one day, and after three days. Treatment: $F(3, 51) = 14.53$, $p < 0.001$.

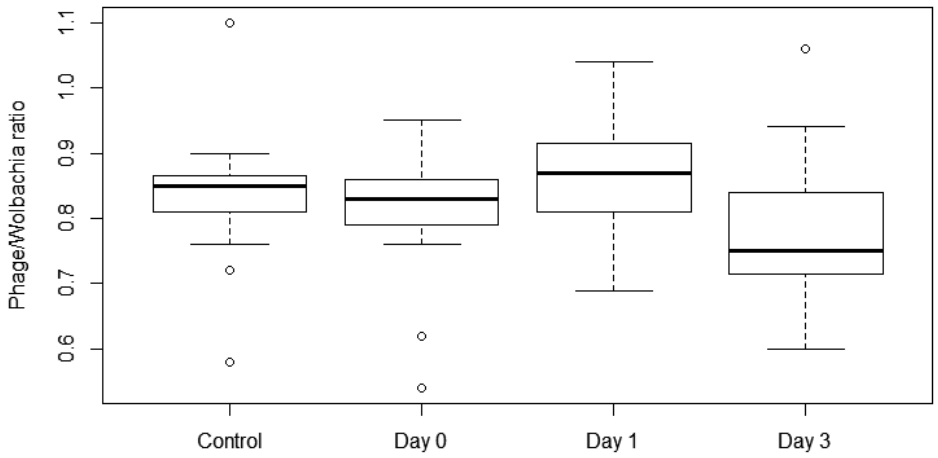


Figure 6: Phage/*Wolbachia* ratio of parthenogenetic *F. candida* injected with *Wolbachia*-containing haemolymph. The boxplots show the median and first and third quartile, with whiskers of 1.5 IQR and outliers indicated by circles. Measurements were taken immediately after the injection (day 0), after one day, and after three days. Treatment: $F(3; 51) = 2,07$; $p = 0,12$.

Curing and re-infection

Of the 30 control springtails, 29 laid two egg clutches and 1 died before it could lay a second clutch (Table 1). In the re-infection treatment, 16 of the 50 springtails died before laying a second egg clutch.

Table 1: Numbers on egg clutches, eggs and juveniles during the curing and reinfection experiment.

		First clutches			**Second clutches**		
	N	Number of clutches	Average number of eggs	Average number of juveniles	Number of clutches	Average number of eggs	Average number of juveniles
Control	30	30	69.5	68.1	29	91.5	89.4
Cured	29	29	40.8	0	29	63.4	0.2
Reinfected	49	49	40.0	0	33	49.2	0.1

The data from the curing and re-infection experiment are displayed in Figures 6 and 7. The second egg clutch was overall significantly larger than the first egg clutch ($F(1; 43) = 29.43$, $p < 0.001$) (Figure 6). Also, a significant difference was found between the three treatments. A Tukey post hoc test showed that the females from the cured and the re-infection group laid significantly smaller egg clutches than the females from the control, $z = -4.92$; $p < 0.001$ and $z = -5.16$; $p < 0.001$.

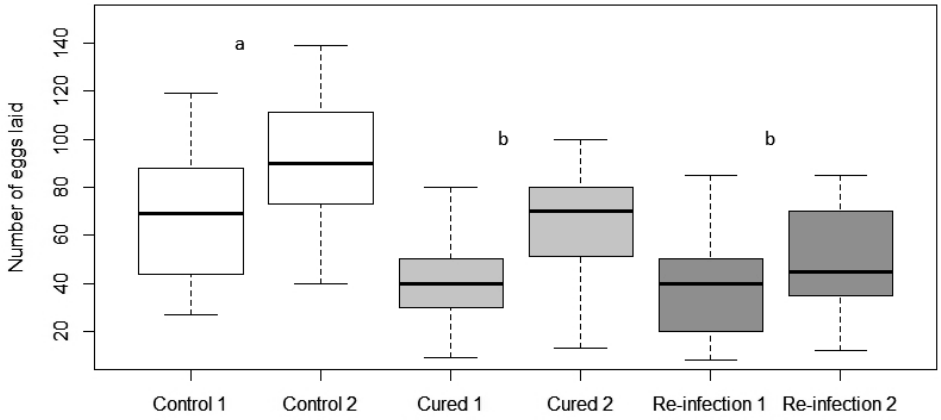


Figure 6: Number of eggs laid by parthenogenetic *Folsomia candida* in the curing and re-infection experiment. The boxplots show the median and first and third quartile, with whiskers of 1.5 IQR. The letters indicate significant differences between the treatment groups. 1 and 2 refer to the first and second egg clutch in the different treatments: control, cured, and re-infected. Clutch 1 vs. 2: $F(1; 43) = 29.43$; $p < 0.001$. Treatment: $F(2, 90) = 42.37$; $p < 0.001$.

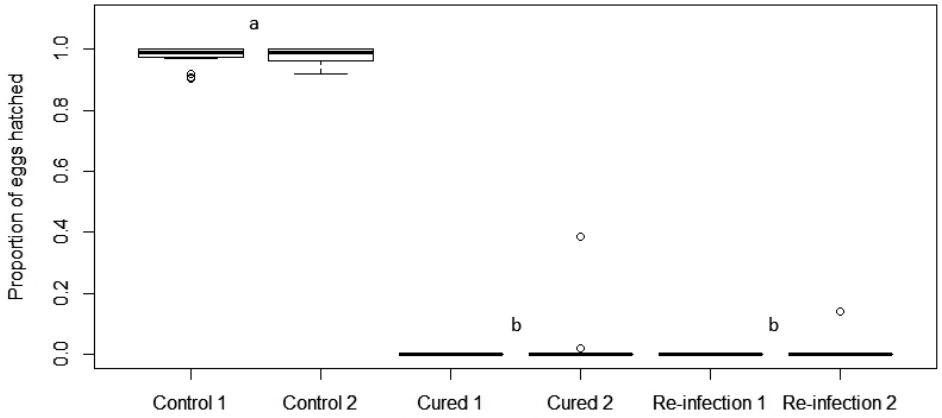


Figure 7: Proportion of eggs laid by parthenogenetic *Folsomia candida* that hatched in the curing and re- infection experiment. The boxplots show the median and first and third quartile, with whiskers of 1.5 IQR. The circles indicate outliers and the letters indicate significant differences between the treatment groups. 1 and 2 indicate the first and second egg clutches of the three treatments: control, cured, and re-infected. Clutch 1 vs. 2: $F(1; 43) = 1.18$; $p = 0.28$. Treatment: $F(2; 90) = 16641.62$; $p < 0.001$.

The percentage of eggs that hatched in both in the first and the second egg clutch of the control was 98% (Figure 7). However, in the cured and the re-infection group, none of the eggs from the first egg clutch hatched. In the second egg clutch, only two of the egg clutches from the cured group had some eggs that hatched (1 egg and 5 eggs), and only one egg clutch from the re-infected group had some hatching eggs (2 eggs). This resulted in a significant difference between the treatments ($F(2; 90) = 16641.62$, $p < 0.001$), with the cured and the re-infected group differing significantly from the control $z = -113.0$; $p < 0.001$ and $z = -116.6$; $p < 0.001$ respectively, as indicated by the Tukey post hoc test. The difference between the clutches was not significant, $F(1; 43) = 1.183$; $p = 0.28$.

Discussion

To gain more insight into the activation of bacteriophage WO when its host *Wolbachia* is under stress, we tested if higher temperatures and transinfection of *Wolbachia* into a new host would increase phage lytic activity. Our results show a significant decrease of *Wolbachia* densities over time in the heat treatment, but no significant difference in phage/*Wolbachia* ratios between the heat-treated springtails and controls. Transinfection into the sexual strain was unsuccessful with almost no *Wolbachia* or phage DNA found, while transinfection into the parthenogenetic strain caused a significant increase in *Wolbachia* densities compared to the control, while the phage/*Wolbachia* ratios stayed unchanged. Finally, the number of eggs per clutch and hatching rate of the *Wolbachia*-cured

springtails were significantly lower than the controls, with a hatching rate as low as zero percent. After discontinuing the antibiotics treatment or re-infecting the springtails with *Wolbachia*, few eggs of the newly laid clutches did succeed in hatching.

By curing *F. candida,* we gained more evidence for the role *Wolbachia* plays in its parthenogenetic reproduction. In concordance with previous studies, antibiotics-treated *F. candida* laid significantly fewer eggs (Pike and Kingcombe, 2009; Timmermans and Ellers, 2009; Giordano *et al.*, 2010). Additionally, the proportion of hatched eggs decreased significantly after the antibiotics treatment with rates dropping from near one to zero for both the first and second clutch. After discontinuing the antibiotics treatment and re-infecting *F. candida* with *Wolbachia* several eggs of the newly laid clutches hatched. No difference in egg hatching between the active increase of *Wolbachia* titer through injection and passive increase through natural recovery after switching to untreated food was found. Nevertheless, our data does support the hypothesis that *Wolbachia* plays a role in the parthenogenetic reproduction of *F. candida*, increasing fecundity and facilitating normal development and egg hatching, which is in accordance with the current knowledge on this topic (Riparbelli, Giordano and Callaini, 2006; Werren, Baldo and Clark, 2008; Pike and Kingcombe, 2009; Timmermans and Ellers, 2009; Giordano *et al.*, 2010).

Our findings did not support the hypothesis that higher temperature would increase the lytic activity of the bacteriophages. This contrasts with previous studies. For example, in spider mites of the species *Tetranychus urticae* the number of bacteriophages to *Wolbachia* ratio increased after being raised under higher temperatures for two generations, suggesting that high temperature increases the lytic activity of the bacteriophages (Lu, Zhang and Hong, 2012). Moreover, an earlier study on the parasitoid wasp *Nasonia vitripennis* showed that the influence of heat on phage density can be life-stage dependent. When wasps are reared from eggs at 30°C instead of 25°C, mean phage densities increase significantly, but not if heat exposure start at the larval, pupal, or adult stages (Bordenstein and Bordenstein, 2011). A possible explanation for why we did not find increased lytic activity could therefore be that we were not able to rear *F. candida* at 30°C from the egg stage onwards. As the eggs would not hatch at 30°C, we started our heat treatment from the juvenile stage. Finding a way to rear *F. candida* from eggs at higher temperatures and do multigenerational heat exposures might still result in activation of the WO phages.

Alternatively, the lack of significant increase in phage/*Wolbachia* ratio in our study could be caused by suppression of the bacteriophages under higher temperatures, keeping them in the lysogenic stage (Burns and Slater, 1982; Tey *et al.*, 2009). A third possibility would be that one of the phages did become lytic, but that this was not distinguishable from the background as we measure a region that is present eight times within the genome, though the expected increase in number

of copies of a lytic phage would be considerably higher. Or finally, it could be that the phages are no longer capable of becoming lytic due to degradation, since transposase activity resulted in the genomic disruption of several loci within the prophage-derived regions (Kampfraath *et al.*, 2019) and as no lytic phage activity has been demonstrated in *w*FolI.

As *F. candida* was exposed to heat stress, the theory that *Wolbachia* titer decreases under high temperatures could be investigated as well. As shown in Figure 4, *Wolbachia* densities indeed decreased gradually over time within our study. A similar decrease under heat stress was also shown in spider mites *T. urticae* that were exposed to heat stress for two generations and the parasitoid wasps *N. vitripennis,* although for the latter it was only observed within the treatment in which eggs were reared at 30°C (Bordenstein and Bordenstein, 2011; Lu, Zhang and Hong, 2012). In both studies they also found a change in the phage titer and concluded that the heat stress activated the phages and consequently decreased the *Wolbachia* population. This contrasts with the results of our study were *Wolbachia* densities decrease, while the phage/*Wolbachia* ratios did not change. Moreover, it implies that the decrease in *Wolbachia* density is in our case a direct effect of the high temperature on the *Wolbachia* cells and not an indirect effect through activation of the phages. Confirming this direct link is of value, as there is an increasing interest into heat resistance of *Wolbachia*, especially for strains that can be used for insect control measures, as this can influence the effectiveness of such biocontrol measures during heat waves (Ross *et al.*, 2017; Burdina *et al.*, 2021; Gu *et al.*, 2022; Ferguson, Ross and van Heerwaarden, 2024).

A second form of stress exposure we imposed upon Wolbachia was transinfection. Infection of a new host exposes the endosymbiont to a novel environment to which it is not yet adapted. Most likely such an event increases stress levels, which may be expected to induce lysis and alter the phage/*Wolbachia* ratio and *Wolbachia* titer. Unfortunately, the transinfections of *Wolbachia* into sexual springtails were unsuccessful. Even immediately after the injection, very low numbers of *Wolbachia* gltA could be detected in the springtails. Thus, the *Wolbachia* concentration within the transferred haemolymph might have been too low to establish a stable infection. However, as shown in figure 6, the number of *Wolbachia* gltA copy numbers present in the parthenogenetic springtails immediately after the injection did increase significantly compared to the parthenogenetic control. This implies that the number of *Wolbachia* present in the injected fluid was sufficient to boost the infection. It remains to be tested however, to what extent *Wolbachia* densities naturally increase over time, as we did not measure infection densities in controls for all time points of the next experiment. Another explanation for not detecting significant numbers of *Wolbachia* gltA after transinfection, could be that the phages were activated immediately and lysed the *Wolbachia* cells. This explanation is not supported, however, by the lack of change observed in phage/*Wolbachia* ratio in the parthenogenetic springtails after the

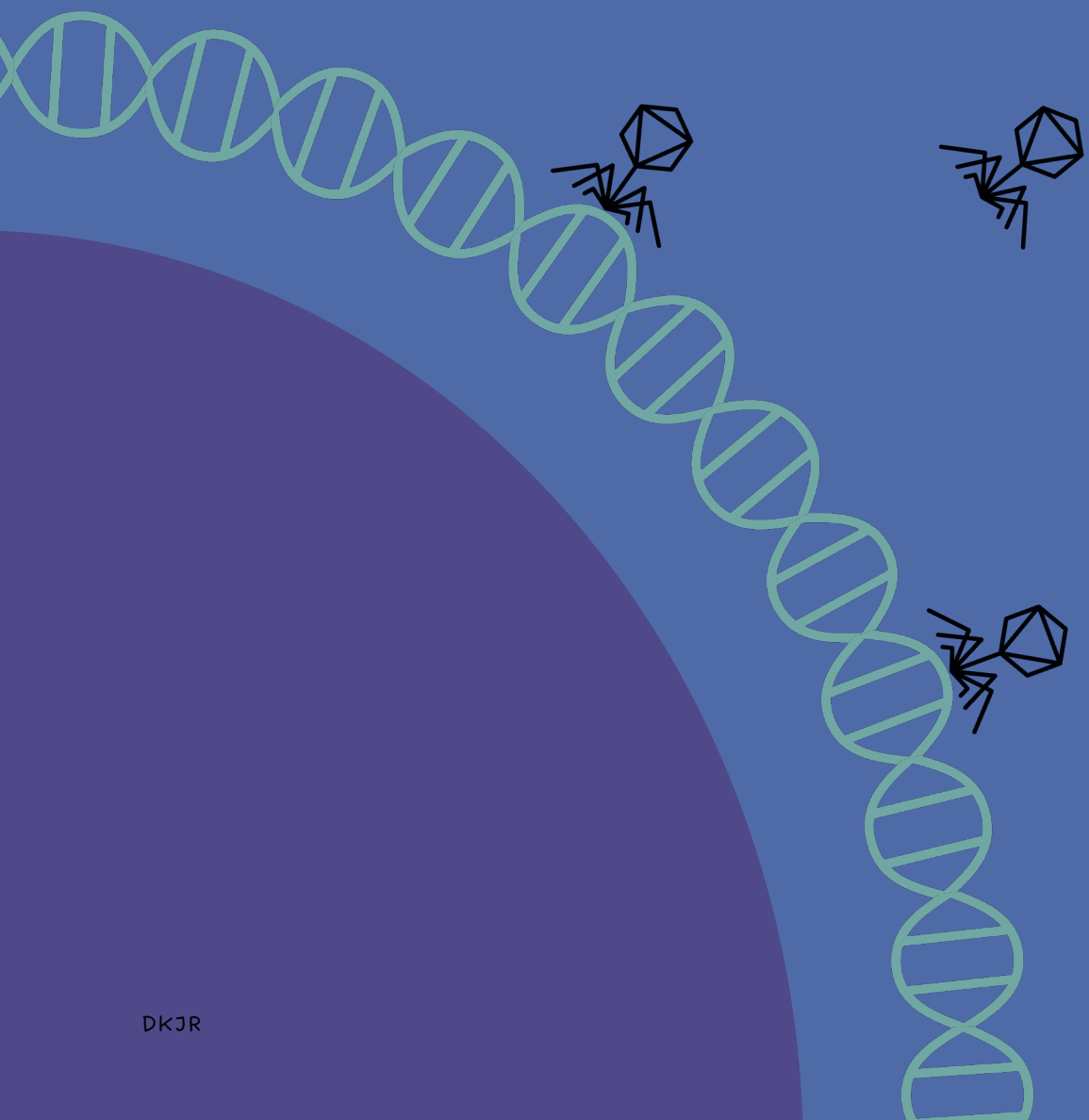
injection. Furthermore, phage activation would have caused phage DNA to be present in the samples, but only in five transinfection attempts in sexual springtails a hint of phage presence was found and all with ct-values were around the detection limit. Consequently, this explanation is improbable as well.

A more probable explanation could be that the immune response of the injected sexual springtails prevented the *Wolbachia* from establishing a stable infection. As the sexual springtails are not the native hosts of *w*FoI, the bacteria probably elicit an immune response (Zug and Hammerstein, 2015). Previous studies in other invertebrates have also shown that infection of a non-native host with *Wolbachia* can induce an immune response. For example, in *Drosophila* cell lines and *Aedes aegypti* mosquitos newly infected with *Wolbachia* immune responses were upregulated (Xi *et al.*, 2008; Kambris and Al, 2009). Moreover, transinfecting *Wolbachia* into the woodlouse *Porcellio d. dilatatus* caused such a strong autophagic response that the animals died prematurely (Le Clec'h *et al.*, 2012). In contrast to the sexual springtails, the parthenogenetic springtails have coevolved with *Wolbachia* and would not have reacted to the familiar bacterium (Zug & Hammerstein 2015). An interesting topic for future research might therefore be whether activation of the immune system after transinfection eradicates the injected *w*FoI cells in sexual *F. candida*.

What would be the implications if indeed the immune system is blocking this transinfection? As the immune system is the foundation of defence against pathogens, this has evolved under strong selection and coevolutionary dynamics. However, many microorganisms that animals encounter are not harmful, and may actually be beneficial. Thus, selection should act on hosts to maintain these associations while preventing exploitation of within-host resources (Gerardo *et al.*, 2020). Invertebrates lack adaptive immunity and use a combination of innate immune functions to differentiate between pathogens and symbionts. For Collembola including *F. candida* several antimicrobial peptide genes have been identified, as part of their innate immune system, and it is hypothesized that this might have led to their long-lasting survival in very microbe-rich habitats, like soil (Faddeeva-Vakhrusheva *et al.*, 2017; Pradhan and Engsontia, 2023). What is still unknown, is whether the sexual and asexual line of *F. candida* contain the same set of antimicrobial peptide genes as this was not a focus within the recent genome comparison (Kampfraath et al. *in prep*). Altogether, the invertebrate innate immune system might be of crucial importance in determining the success of endosymbiont transinfection.

In summary, our results suggest that high temperature has a direct negative effect on *w*FoI rather than an indirect effect through activation of bacteriophages, with the most probable explanation being that the bacteriophages are no longer able to become lytic. Furthermore, transinfection of *Wolbachia* into the sexual lineage of *F. candida* was unsuccessful, which may have been caused by an immune response in the sexual springtails.

CHAPTER 4



DKJR

Genome expansion of an obligate parthenogenesis-associated *Wolbachia* poses an exception to the symbiont reduction model

A.A. Kampfraath, L. Klasson, S.Y. Anvar, R.H.A.M. Vossen, D. Roelofs, K. Kraaijeveld & J. Ellers

Status: published in BMC Genomics

Background: Theory predicts that dependency within host-endosymbiont interactions results in endosymbiont genome size reduction. Unexpectedly, the largest *Wolbachia* genome was found in the obligate, parthenogenesis-associated *w*Fol. In this study, we investigate possible processes underlying this genome expansion by comparing a re-annotated *w*Fol genome to other *Wolbachia* genomes. In addition, we also search for candidate genes related to parthenogenesis induction (PI).
Results: Within *w*Fol, we found five phage WO regions representing 25.4% of the complete genome, few pseudogenized genes, and an expansion of DNA-repair genes in comparison to other *Wolbachia*. These signs of genome conservation were mirrored in the *w*Fol host, the springtail *F. candida*, which also had an expanded DNA-repair gene family and many horizontally transferred genes. Across all *Wolbachia* genomes, there was a strong correlation between gene numbers of *Wolbachia* strains and their hosts. In order to identify genes with a potential link to PI, we assembled the genome of an additional PI strain, *w*Lcla. Comparisons between four PI *Wolbachia,* including *w*Fol and *w*Lcla, and fourteen non-PI *Wolbachia* yielded a small set of potential candidate genes for further investigation.
Conclusions: The strong similarities in genome content of *w*Fol and its host, as well as the correlation between host and *Wolbachia* gene numbers suggest that there may be some form of convergent evolution between endosymbiont and host genomes. If such convergent evolution would be strong enough to overcome the evolutionary forces causing genome reduction, it would enable expanded genomes within long-term obligate endosymbionts.

Key words: *Folsomia candida*, *Leptopilina clavipes, w*Fol, *w*Lcla, parthenogenesis induction

Background

Endosymbiotic bacteria are found in the cells of many eukaryotic species, where they affect major processes such as host reproduction (Stouthamer, Breeuwer and Hurst, 1999; Cordaux, Bouchon and Grève, 2011), defence against pathogens (Scarborough, Ferrari and Godfray, 2005) and development (Dedeine *et al.*, 2001; Gill, Darby and Makepeace, 2014). Endosymbionts are dependent on their host for proliferation, whereas host are not necessarily dependent on their endosymbiont and can often reproduce and survive in their absence. However, there are ample examples of obligate endosymbionts, e.g. *Buchnera* in aphids (Douglas, 1998; Shigenobu *et al.*, 2000), *Wigglesworthia* in the tsetseflies (Akman *et al.*, 2002), and *Wolbachia* in parasitic filarial nematodes (Foster *et al.*, 2005; Darby *et al.*, 2012). In these cases neither host nor endosymbiont are viable without the other, and these associations are usually characterised by a long evolutionary history, nutritional or developmental dependency, and vertical transmission of the symbiont (Moran and Bennett, 2014).

Current theory predicts that mutual dependency between host and endosymbiont leads to a reduction in genome size of endosymbionts (Moran, McCutcheon and Nakabachi, 2008; McCutcheon and Moran, 2012b). Different genetic mechanisms have been implied as the driving forces behind these reductions. First, co-adaptation between the symbiont and their host may result in redundancy of certain symbiont functions releasing genes from selective constraints, causing them to decay and eventually to disappear (Dale and Moran, 2006; Ellers *et al.*, 2012). Second, small population sizes and limited opportunity for horizontal gene exchange increases genetic drift to a level that purifying selection cannot overcome, which leads to deleterious mutations that cause pseudogenization of mildly advantageous genes and the eventual removal of those genes due to the inherent deletion bias found for bacterial genomes (Moran, 1996; Mira, Ochman and Moran, 2001; McCutcheon and Moran, 2012). In addition, multiple transitional events seem to pinpoint moments of sever genome reduction (Lo, Huang and Kuo, 2016). These include becoming host bound, moving into a specialized host cell and being vertically transmitted through host generations. Intervals between different transitional events seem prone to different sets of selection pressures that influence genome size. For example, the negative correlation between host dependence and symbiont genome size only holds for vertically transmitted endosymbionts (Fisher *et al.*, 2017). However, there is still a large amount of variation in the genome size of vertically transmitted endosymbionts that are not residing in specialized host cells, which is difficult to explain in the context of current theory (Lo, Huang and Kuo, 2016). Examining endosymbiont genomes in this stage will be needed to better understand all factors that influence the genome reduction during the evolution of endosymbionts.

Wolbachia is one of the most widespread endosymbionts (Hilgenboecker *et al.*, 2008; Zug and Hammerstein, 2012) and known for its variety of interactions with its host, including male killing, feminization, cytoplasmic incompatibility (CI), parthenogenesis-induction (PI) and provisional mutualisms (Werren, 1997; Merçot and Poinsot, 2009). Plus, all known *Wolbachia* strains are in a similar transitional stage according to the classification of Lo et al. (2016), i.e. they are mainly transmitted vertically and do not reside in specialized structures. They are thus predicted to vary in genome size depending on the host dependency (Lo, Huang and Kuo, 2016). This prediction seems to hold for most of the *Wolbachia* strains, with reduced genomes that lack mobile elements found in obligate mutualistic strains and larger genomes in facultative reproduction-manipulating strains. However, recently a *Wolbachia* genome was sequenced that poses an exception to this pattern. The *Wolbachia* from the parthenogenetic collembolan host *Folsomia candida* has the largest *Wolbachia* genome sequenced so far (Faddeeva-Vakhrusheva *et al.*, 2017), yet it is obligate to its host. Since, elimination of *Wolbachia* by heat or antibiotics, renders the host's eggs non-viable (Pike and Kingcombe, 2009; Timmermans and Ellers, 2009). In addition, the *w*Fol genome is the first fully sequenced genome from supergroup E that is positioned to be a sister group to supergroup A, B, C, D, F and H, and shares a more ancestral common ancestor with supergroup L (Gerth *et al.*, 2014; Brown *et al.*, 2016). All complete genomes from members of these supergroups have smaller genomes than *w*Fol, which suggests that this obligate *Wolbachia* strain showed genome expansion rather than the reduction that would be predicted by the symbiont genome reduction model.

Furthermore, the *w*Fol genome could contribute to understanding the genes and mechanisms underlying PI, as *w*Fol has been suggested to cause parthenogenesis in its host. All parthenogenetic *F. candida* lines carry *Wolbachia* whereas the sexually reproducing lines do not (Czarnetzki and Tebbe, 2004; Frati *et al.*, 2004; Timmermans and Ellers, 2009). Unequivocal proof of PI would require curing of a parthenogenetic strain to produce males, but since *w*Fol induces parthenogenesis in addition to promoting host egg development, this experiment is not possible in the diplo-diploid *F. candida*. In several Hymenopteran species, *Wolbachia* has been shown to induce diploidy and feminization resulting in parthenogenetic reproduction (Pannebakker *et al.*, 2004; Ma *et al.*, 2015) and several draft genomes of PI-*Wolbachia* from Hymenopteran host are available: *w*Tpre from *Trichogramma pretiosum* and *w*Uni from *Muscidifurax uniraptor* (Klasson *et al.*, 2009; Lindsey *et al.*, 2016; Newton *et al.*, 2016). Building a larger database of *Wolbachia* genomes associated with PI might shed more light on the genes involved in PI, similar to recent finding for CI (Le Page *et al.*, 2017).

In this study, we set out to explain the factors that contribute to the expansion in the *w*Fol genome, which can help further understand the genomic evolution within endosymbionts. Therefore, we updated and re-annotated the

*w*Fol genome and compared it to a diverse set of high quality *Wolbachia* genomes. In addition, we searched for genes shared by *Wolbachia* associated with PI that could aid in understanding the mechanism behind this manipulation. To this end, we assembled a draft genome of an additional PI strain, *w*Lcla, that has been shown to cause diploidy restoration in the parasitoid wasp *Leptopilina clavipes* (Pannebakker *et al.*, 2004). Finding the genes underlying PI in *Wolbachia* might elucidate the mechanism behind it and could resolve the debate on the nature of *w*Fol host interaction.

Results

Genome assemblies of wFol and wLcla

The short read corrected *w*Fol genome consisted of one contig with a total length of 1,801,626 bp and a GC-content of 34.35%. This was 43 bp longer than the initial assembly and the GC-content was a half percent higher (Faddeeva-Vakhrusheva *et al.*, 2017). The draft genome of *w*Lcla was assembled in 46 contigs with a total length of 1,150,755 bp and a GC content of 34.11%. Half of the assembly was contained in 9 contigs (L50) larger than 43,523 bp (N50). For a *Wolbachia* genome this is relatively small, but very similar in size to the other PI strain from supergroup B (*w*Tpre) (Lindsey *et al.*, 2016). However, *w*Tpre is in one scaffold that contains a total gap length of 16,680 bp and *w*Lcla is in 46 pieces, thus we still miss information for both of these genomes. Hence, the current assembly of *w*Lcla probably represents the major part the complete genome, but it will presumably increase slightly in size upon completion.

To evaluate the quality and completeness of our annotations, we run the BUSCO pipeline based on 148 essential bacterial genes on the annotations of *w*Fol, *w*Lcla and an additional 16 *Wolbachia* annotations to compare with (see Additional file 1). The *w*Fol genome contained the most complete set of essential genes of the 18 *Wolbachia* analysed. It contained the complete sequences of 92.6% of the essential genes, including two duplications, and two fragmented. Only nine genes were missing, which was the lowest number of all *Wolbachia* analysed. The BUSCO gene set to which it was compared is a curated set of genes that are essential to most bacteria. However, in *Wolbachia* even complete genomes miss 9 to 25 genes from the BUSCO set, indicating that these genes probably are not missing from the assemblies and annotations but because they have become redundant for *Wolbachia* due to its endosymbiotic lifestyle. In the draft genome of *w*Lcla we found 133 complete (89.8 %), two fragmented and 11 missing genes, which is very similar to most of the other *Wolbachia* genomes.

Table 1: Genomic characteristics of 18 *Wolbachia* strains used for comparisons.

Strain	Size (Mbp)	Contigs	CDS	Pseudo	Transposases	ANK	Repair	Resolvases
*w*Au	1.27	1	1204	62	119	39	18	4
*w*Bm	1.08	1	805	98	0	5	12	0
*w*Bol1-b	1.38	144	1139	162	11	2	23	2
*w*Cle	1.25	1	1216	NA	244	49	12	0
*w*Fol	1.80	1	1472	44	124	96	34	19
*w*Ha	1.30	1	1009	96	19	29	17	5
*w*Lcla	1.15	46	880	194	50	22	12	1
*w*Mel	1.27	1	1195	74	39	23	21	2
*w*No	1.30	1	1040	90	19	48	18	2
*w*Oo	0.96	1	647	195	0	3	7	0
*w*Ov	0.96	1	642	42	0	0	7	0
*w*Pip_Pel	1.48	2	1275	110	62	42	20	5
*w*Ppe	0.98	36	851	62	1	0	13	1
*w*Rec	1.12	43	902	127	26	2	13	1
*w*Ri	1.45	1	1150	114	67	29	18	4
*w*Tpre	1.13	9	1405	NA	53	54	16	2
*w*Uni	1.05	130	1174	NA	67	18	17	0
*w*VulC	1.66	10	1293	255	20	27	19	9

Column label abbreviations: CDS stands for coding sequences and ANK for ankyrin repeat containing protein. Annotations of *w*Cle, *w*Tpre and *w*Uni did not specify pseudogenes; for these strains this category is therefore labelled as not available (NA).

Annotation of wFol and wLcla

Annotating the improved *w*Fol genome uncovered a total of 1472 protein coding sequences (CDS), which is the largest number of CDS for any *Wolbachia* genome sequenced so far. This is not surprising given that it is the largest *Wolbachia* genome, and genome size and gene number correlate well in bacteria. However, a remarkably low number of pseudogenes were found in its genome. Only 2.8 % of the genes appeared to be pseudogenized, while the pseudogene content in the other *Wolbachia* genomes ranged from 4.7 % to 22.1 %. In addition, we observed a high number of transposases, DNA-repair related genes and ankyrin repeat containing proteins (ANKs; see Table 1). The number of ANKs (96) and DNA-repair related genes (34) in *w*Fol outnumbered the quantity found in any other *Wolbachia* and the number of transposases is second only to the number found in *w*Cle.

In addition, we found large numbers of phage genes, which were concentrated in five regions of phage origin (RPO) that were labelled WOFol1 to 3 and WO-like island 1 and 2 (see Additional file 2: Supplementary table 1). With a combined size of 458,452 base pairs, they took up 25.4% of the total genome size

of *w*Fol, the largest amount in any *Wolbachia*. Remarkably, pseudogenes and the three before-mentioned overrepresented groups of genes (ANKs, DNA-repair related and transposases) were all unequally distributed between the phage regions and rest of the genome (see Figure 1A). Of all groups more than the expected 25.4% was present within the RPOs, finding 29 out of 44 pseudogenes (binomial-test, R, p<0.001), 23 out of 34 DNA-repair related genes (binomial-test, R, p<0.001), and 42 out of 96 ANKs within them (binomial-test, R, p<0.001). The distribution of transposases over the phage regions and rest of the genome was not different from a random expectation, since 26.6 % of them were found in the RPOs (binomial-test, R, p=0.75). However, when focussing on individual transposase families, they did show unequal distribution patterns (see Figure 1B). While the IS4, IS110 and IS5 families were proportionally distributed over the whole genome, the PD-(D/E)XK family was more common in the RPOs than expected (binomial-test, R, p<0.001). In addition, two of the other five families exclusive to either the RPOs or to the chromosomal part of the genome also showed a significantly skewed distribution (IS481: binomial-test, R, p<0.01; IS5, IS903: binomial-test, R, p<0.05).

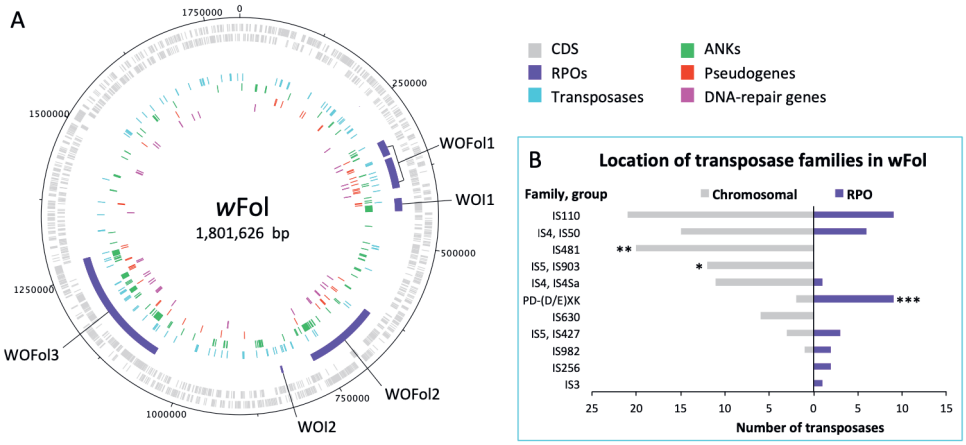


Figure 1: Distribution of genomic elements of interest in the *w*Fol genome. A) Circular map of *w*Fol displaying the distribution of genes of interest between regions of phage origin (RPOs) and chromosomal parts of the genome. Each ring gives locations of certain genes or genomic elements. From outer to inner ring: CDS forward strand, CDS reverse strand, RPO, transposases, ankyrin repeat containing genes (ANKs), pseudogenes and DNA-repair related genes. RPOs are separately labelled: WO regions with WOFol1 to 3 and WO-like islands with WOI1 and 2. B) Double side histogram that presents the distribution of transposase groups between the chromosomal and the RPO. Distributions were analysed with a binomial test in R, asterisks correlate to significant levels * to p > 0.05, ** to p > 0.01 and *** to p > 0.001.

The annotation of *w*Lcla uncovered 880 CDS and 194 pseudogenes (see Table 1). Thus, while it is similar in size to *w*Tpre, it contains 525 fewer CDS and 194 more

pseudogenes. This is probably an effect of differences in annotation style, as no pseudogenes were annotated for *w*Tpre, but many truncated genes were found (Lindsey *et al.*, 2016) that were therefore counted as one or two CDS instead of one pseudogene. Finally, *w*Lcla contained 4 phage-related genes as well as 7 phage-related pseudogenes, thus showing signatures of remnant phages. However, no conserved RPOs could be detected within the assembled contigs.

Phage annotation and synteny

As the RPOs take up more than a quarter of the *w*Fol genome we put extra effort in their annotation. Kent et al. (2011) compared several WO phages and found that they are composed of multiple modules of genes that link to certain functions. We also found such modules in the WOFol phages, which included patatin-like phospholipases and baseplate, head and tail modules (see Figure 2A). However, not every module was present in all three WOFol regions, and WOFol2 and 3 contained many replicated modules. With a size of respectively 132.270 and 215.744 WOFol2 and 3 were the largest WO regions found in any *Wolbachia* (see Additional file 2: Supplementary table 1). Surprisingly, all three WOFol regions contained the tail module, which is often lost in other WO (Kent *et al.*, 2011). In addition, WOFol1 was missing the head and baseplate modules, but did contain the conserved WD0611-WD0621 cluster that is also found in several other WO phages (Bordenstein and Bordenstein, 2016). All three WOFol regions also contained the recently described eukaryotic association module (EAM) (Bordenstein and Bordenstein, 2016) and within them 6 of the WO Latrotoxin CTD proteins. Further, the RPO of *w*Fol contained 19 resolvases, which is at least two times the number of resolvases found in any other *Wolbachia* genome (see Table 1). Resolvases are often connected to phage integration at specific sites, while integration by transposases is less site-specific (Campbell, 1992). Finally, WOFol 2 and 3 also harbored a bacteriophage abortive infection system, the AbiEi and AbiEii antitoxin-toxin complex, which has never been found in WO before. This system can provide phage protection at the population level by killing its host when infected by a new phage (Dy *et al.*, 2014).

The large number of RPOs in *w*Fol led to the question if all these regions arose from a single ancestral phage that duplicated itself or, whether multiple phages infected the *w*Fol genome. Therefore, we searched for collinear blocks in the RPO and found them both between the three WOFol regions and within WOFol2 and 3 (see Figure 2A-B). However, based on the mosaic structure of the collinear blocks there does not seem to be a complete duplication of one of the WO phages.

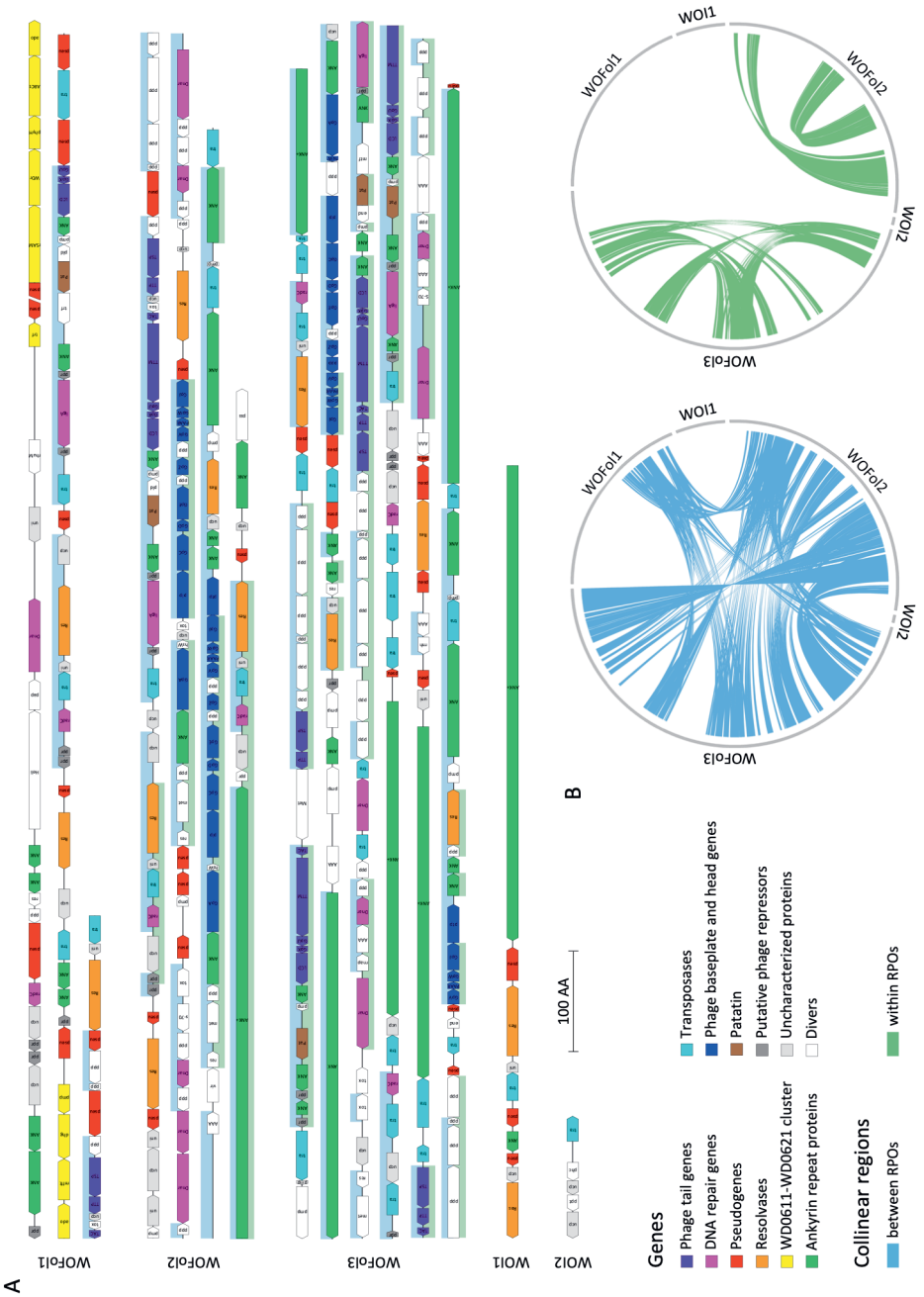


Figure 2: Detailed graphical overview of phage WO regions in *w*Fol. A) Graphical view of the RPO annotations of *w*Fol. Arrow blocks depict genes, upper marked regions are within collinear blocks between and lower marked regions within RPOs. A list of the microscopic abbreviations of gene annotations within the arrow blocks can be found in Additional file 6. B) Circular maps of collinear blocks between the RPOs of *w*Fol on the left (blue) and within (green) on the right.

Orthologue identification and orthogroup expansion

Orthofinder was run to cluster the complete set of 19,303 protein-coding sequences (CDS) of the 18 *Wolbachia* into orthogroups resulting in 1,239 orthogroups containing 18,480 CDS (95.7%) (see Additional file 3: sheet Overall statistics). Of these, 460 were present in all *Wolbachia* genomes, including 414 that only contained single copy genes (see Additional file 3: sheet Orthogroups shared by all). These 460 genes can be considered the *Wolbachia* core genome, which is within the range of the 489 orthogroups found by Brown et al. (2016), when comparing members of supergroups A-D, F and L. For *w*Fol, 1,346 of the 1,472 CDS were grouped into 936 of the orthogroups (see Additional file 3: sheet All orthogroups). Thus, 126 CDS remained unassigned and were therefore considered to be unique to *w*Fol (see Additional file 3: sheet Unassigned genes). *w*Tpre was the only genome containing more unique CDS. This was most likely due to the many truncated/pseudogenes in this genome annotation (Lindsey *et al.*, 2016), which probably have ended up in separate groups because of Orthofinder's algorithm sensitivity for gene length (Emms and Kelly, 2015), thus creating false orthogroups of truncated genes. Most of the unassigned genes from *w*Fol were hypothetical proteins (76) and putative membrane proteins (27), but 23 of these genes could be annotated in more detail (see Additional file 3: sheet Unassigned genes). Moreover, 35 (27.8%) of the unassigned CDS were located within the RPOs, which include the AbiEi and AbiEii antitoxin-toxin complex mentioned above (Additional file 4). For the *w*Lcla draft genome, 875 of their 879 CDS were grouped into 817 orthogroups, while only 4 remained unassigned. Three of these were annotated as hypothetical proteins and one was a transposase (see Additional file 3: sheet Unassigned genes). Therefore, no distinct functions could be linked to these genes.

To define orthogroup expansions and contractions, Z-scores were calculated for each orthogroup (see Figure 3 and Additional file 3: sheet all orthogroups). Z-scores measure the deviation of the number of genes of one strain from the average number of genes per strain in an orthogroup. Orthogroup expansions had more genes (Z-score 2 or higher) and contractions fewer (Z-score -2 or lower).

The three genomes with the highest number of gene family expansions were all linked to PI (*w*Tpre, *w*Uni and *w*Fol). Unfortunately, no orthogroup was expanded in all three *Wolbachia* genomes. Focussing on the *w*Fol genome we found 72 expanded and 10 contracted orthogroups (see Additional file 3: sheet Expanded or contracted in *w*Fol). All 10 contracted orthogroups were genes completely missing in *w*Fol. The 72 expanded orthogroups contained 371 genes of which more than half within the RPOs (194 genes within 21 orthogroups). In line with this, 89 genes were phage-related genes, pressing the notion that the expansion of *w*Fol is for a large part due to an increase of phage genes. In addition, more than half (190)

of the genes were transposases (93), ANKs (74) and DNA-repair related (21), proving that these groups are overrepresented in this genome.

Not surprisingly, considering its small size the *w*Lcla genome contained only 14 expanded orthogroups with 26 genes in total (see Additional file 3: sheet Expanded or contracted in *w*Lcla). Eight of these genes were within one orthogroup of IS110 family transposases, while all other groups consisted of 1 gene missing in all other *Wolbachia* or 2 duplicate genes in *w*Lcla. Most of these other groups consisted of uncharacterized proteins without a clear function and the genes with annotation had diverse functions in several biological processes. The *w*Lcla genome contained only five significantly contracted orthogroups, which is surprising given the large number of pseudogenes. This indicates that many of the pseudogenized genes of *w*Lcla were also degraded in other *Wolbachia.*

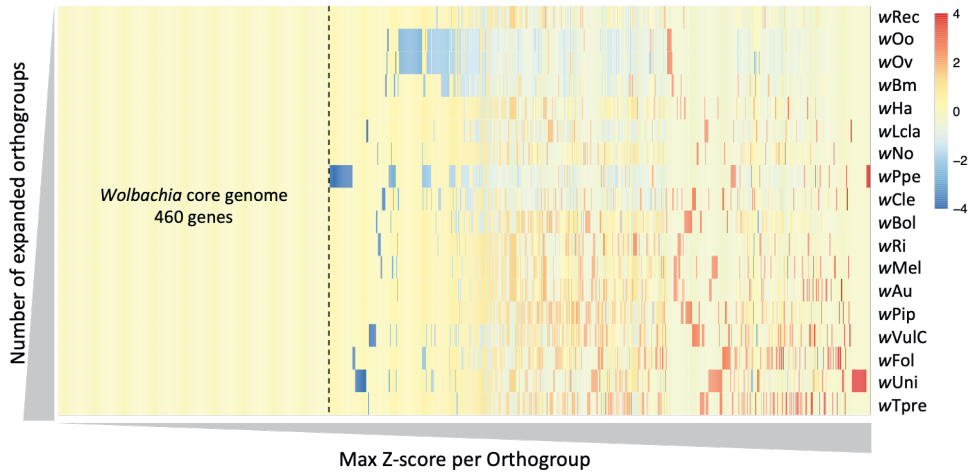


Figure 3: Expanded and contracted orthogroups in different *Wolbachia* strains. Heatmap of the Z-scores of the 18 *Wolbachia* strains per orthogroup that indicates expansions in red and contractions in blue. The rectangle separated by a dashed line represents the 460 core genome genes without variation in Z-score. *Wolbachia* strains are ordered by the number of expanded orthogroups from the least on top till the most at the bottom and all orthogroups are sorted on their max Z-score from lowest at the left to highest at the right.

Parthenogenesis induction genes

Clustering all genes into orthogroups also allowed us to search for orthogroups whose absence or presence was associated with the four PI-*Wolbachia* strains (*w*Fol, *w*Lcla, *w*Tpre and *w*Uni). We found no genes that were exclusively missing from the four PI-*Wolbachia*. There was a large number of shared orthogroups (684) between the four genomes, which obviously also included the 460 core genome orthogroups (see Figure 4A). However, there were no orthogroups that were exclusively present in all four or three out of the four parthenogenesis-associated

Wolbachia genomes (see Figure 4B). Only between pairs of the PI-*Wolbachia* 11 exclusive orthogroups existed (see Additional file 3: sheet Exclusive or mainly in PI). Five of those contained only uncharacterised proteins and another two contained only transposases. The remaining four consisted of: 1) a putative phage protein shared by *w*Uni and *w*Fol, 2) a protein with a magnesium transported domain shared between *w*Tpre and *w*Uni and 3) a cluster of two genes shared between *w*Tpre and *w*Lcla consisted of a protein with a partial fungal domain of unknown function and a protein that contains several leucine-rich repeats. These leucine-rich repeats are protein binding domains that are involved in a wide range of biological processes (Rothberg *et al.*, 1990). Furthermore, we also searched for orthogroups that were present in at least three of the four PI strains and in a maximum of two other *Wolbachia* strains. This yielded two orthogroups with single copy uncharacterized genes, represented in *w*Fol by wFol_04740 and wFol_12640, and both shared by *w*Fol, *w*Lcla, and *w*Tpre (see Additional file 3: sheet Exclusive or mainly in PI). Absence of these genes in *w*Uni strain might be inaccurate, as the still quite fragmented assembly is possibly hampering gene annotation (Table 1). One orthogroup was also found in the male-killing strain *w*Bolb-1 and the other orthogroup in both *w*Bolb-1 and CI-strain *w*Pip_Pel. Based on their annotation neither of the examined genes could be directly linked to parthenogenesis induction.

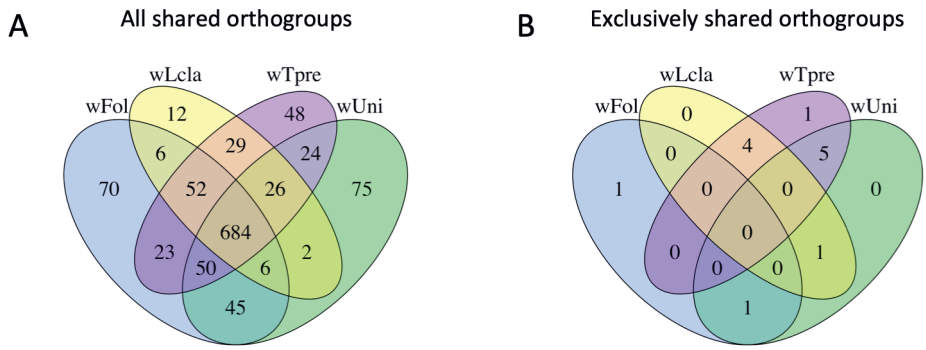


Figure 4: Orthogroups shared by PI *Wolbachia***.** Venn diagrams representing A) a distribution of all orthogroups shared between the PI associated *Wolbachia*, including the orthogroups shared with the other 14 *Wolbachia* and B) a distribution of all orthogroups exclusively present in the PI associated but not in the other 14 *Wolbachia*.

Whole genome correlations

We found a positive correlation between *Wolbachia* genome size and the amount of phage DNA it contained (see Figure 5A). This indicates that a significant portion of the variation in genome size was accounted for by the RPOs. The genomes of

Wolbachia strains that are essential to their host occurred at both ends of the size spectrum, with most having small genomes and little phage DNA and *w*Fol having a large, phage-rich genome. In addition, we also found a positive correlation between the number of DNA-repair related genes and the genome size of *Wolbachia* (see Figure 5B).

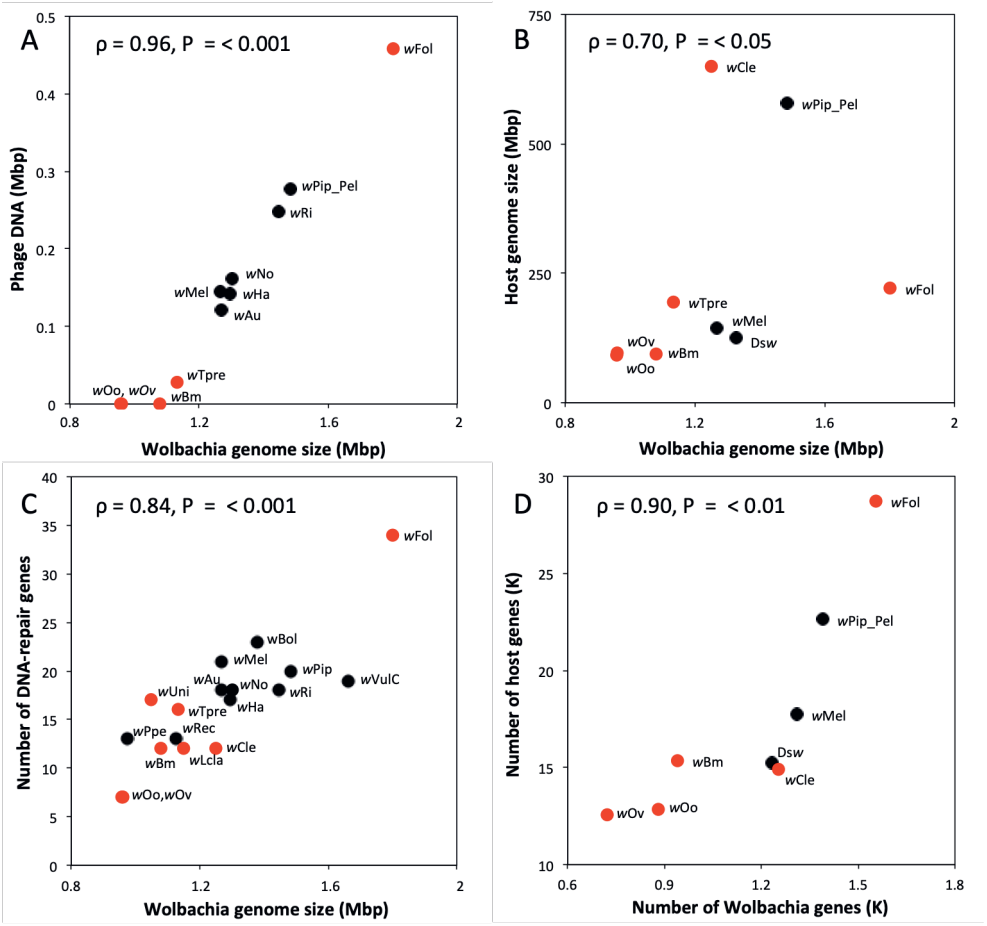


Figure 5: Multiple positive correlations associated with genome evolution in *Wolbachia*. Correlations between A) genome size and amount of integrated phage DNA both in number of basepairs, B) host and *Wolbachia* genome size in mega basepairs, C) number of DNA-repair related genes and *Wolbachia* genome size and D) host and *Wolbachia* number of genes. Red data points indicate obligate transitional and black facultative endosymbionts; all points are labelled with strain names and for the correlations with the host the four strains residing in *Drosophila simulans* were averaged and labeled as Dsw. Spearman's rank correlations and corresponding P-values calculated in R are shown in the top corner of every plot.

Host genome size and gene number were also correlated to *Wolbachia* genome size and gene number, respectively (see Figure 5C-D), with gene content being more strongly correlated. Moreover, the correlation between genome sizes still holds when RPO are excluded (ρ = 0.86, P = 0.01; see Additional file 5), although the deviating strain *w*Cle needed to be excluded as the RPO could not be properly annotated and therefore their size could not be determined. These correlations suggest that the genome size and number of genes of a *Wolbachia* are not independent from the size and number of genes of its host genome.

Discussion

The *w*Fol genome is the largest *Wolbachia* genome to date, and our results suggest that its genome size has increased due to the integration of several phages, as phage regions take up more than a quarter of the total *w*Fol genome. This is supported by an expansion of 21 orthogroups containing phage genes. These regions of phage origin (RPOs) had higher numbers of ankyrin repeat containing genes (ANKs), DNA-repair related genes, pseudogenes and PD-(D/E)XK transposases. In addition, they contained a bacteriophage abortive infection system new to *Wolbachia*. Comparing *w*Fol, *w*Lcla and two other PI-associated strains to a diverse set of 14 *Wolbachia* genomes did not elucidate a set of genes unique to the four PI strains. However, there were unique sets of genes between some pairs of the PI associated strains. Finally, we found strong correlations between *Wolbachia* and host gene numbers, suggesting that there might be convergent evolution between *Wolbachia* and their host genomes.

Genome expansion in wFol

The first question we set out to answer was why the genome of the obligate *Wolbachia* endosymbiont of the parthenogenetic springtail *F. candida* (*w*Fol) has expanded, while current theory predicts an obligate relationship to lead to a reduced genome size (McCutcheon and Moran, 2012; Lo, Huang and Kuo, 2016). Obligate endosymbionts usually have a small genome size and contain few repeated and mobile elements such as transposases (Cerveau *et al.*, 2011), ANKs (Iturbe-Ormaetxe *et al.*, 2005; Siozios *et al.*, 2013) and RPOs (Masui *et al.*, 2000; Kent and Bordenstein, 2010). Genome reduction of this kind is also evident in the genomes of obligate *Wolbachia* strains in filarial nematodes (*w*Oo, *w*Ov and *w*Bm), which are the smallest complete *Wolbachia* genomes, containing no or very few transposable elements and phage derived genes (Foster *et al.*, 2005; Darby *et al.*, 2012; Comandatore *et al.*, 2015). In contrast, the genome of the obligate *w*Fol strain is the largest complete *Wolbachia* genome discovered to date with a length of 1,801,626 base pairs (bp) (Faddeeva-Vakhrusheva *et al.*, 2017). Moreover, upon annotating this unusually large genome meticulously, we found that it is highly

enriched in repeated and mobile elements and that the RPOs take up more than a quarter of its genome. These regions contain many of the ANKs and transposases that are enriching this genome and have a big influence on the genomic structure and size of *w*Fol.

Many expanded orthogroups were also found to be located in the RPO, including a set of genes connected to DNA repair. Integrated phage genes within endosymbionts can still be expressed and are known to influence bacterial and host processes (Weldon, Strand and Oliver, 2013; Le Page *et al.*, 2017). Therefore, the functions of these integrated genes may affect the performance of the tri-partite symbiosis. Typically, in obligate endosymbiont genomes a depletion of DNA repair genes is seen, and the loss of these genes would result in a higher effective mutation rate (Moran, McCutcheon and Nakabachi, 2008). Thus, the gain in DNA repair genes in the *w*Fol genome can be expected to lower effective mutation rate, resulting in a better-conserved genome with fewer deletions and fewer pseudogenized genes.

We also found many unique genes present in *w*Fol. Although RPOs are known to be a source of new genes (Klasson *et al.*, 2008; Siozios *et al.*, 2013; Bordenstein and Bordenstein, 2016), in *w*Fol unique genes were evenly distributed over the genome. A possible explanation for the large number of novel genes could be that *w*Fol is the first annotated *Wolbachia* genome of supergroup E, while many of the other supergroups were represented with more genomes in our analysis. Therefore, this group of unique genes might not just represent genes unique to *w*Fol but also include genes that are specific to the entire E supergroup. Nevertheless, the presence of this many unique genes indicates that either *w*Fol specifically or all members of supergroup E are able to acquire new genes or preserve existing ones more easily compared to other *Wolbachia*.

Another strong signature of gene conservation is the low number of pseudogenes found in *w*Fol. Most of the pseudogenes that were present were located within the RPO, suggesting a less stringent conservation of phage-derived genes. This is in line with what is seen in other *Wolbachia* where the phage regions are often more prone to degradation (Metcalf *et al.*, 2014). However, the rather high share of intact phage genes within *w*Fol, suggest that even these regions with higher degradation rates are still being maintained at an elevated rate. The combined results on phage preservation, the low number of pseudogenes, the many DNA repair genes, and the large number of unique genes consistently point towards a genome shaped by gene preservation. This is inconsistent with expectations based on the symbiont genome reduction model.

At first glance, the *w*Fol genome with its relatively large size and many mobile elements resembles a recently host-restricted symbiont. However, these symbionts are generally facultative for their host, harbour many pseudogenes and suffer large and small deletions, as their genomes change rapidly (McCutcheon and Moran, 2012; Lo, Huang and Kuo, 2016). By contrast, *w*Fol is obligate to its host,

possesses only few pseudogenes, and shows no signs of large deletions. Furthermore, the start of the symbiosis between *w*Fol and its host probably coincided with the host becoming parthenogenetic. Although, this split has not been properly dated yet, genetic evidence shows considerable differentiation between sexual and parthenogenetic lines of *Folsomia candida* (Frati *et al.*, 2004), suggesting this was not a recent event. Thus, *w*Fol does not fit the description of a reduced long-term obligate symbiont genome, nor a recently host-restricted symbiont genome. Rather, it seems to fall into a new category of a long-term obligate symbiont that was able to avoid genome degradation.

This leads to the question: what has caused the genome preservation in *w*Fol compared to the other *Wolbachia*, including the other PI and obligate strains? These obligate strains are just like *w*Fol vertically transmitted within hosts that are dependent on their *Wolbachia*, thus there must be another explanation for the differences in genome maintenance than level of interdependence with their host or form of transmission. Interestingly, there are unique features to the genome of *F. candida*, which mirror the genomic patterns of its symbiont. Its genome has with 28,734 genes the most gene rich genome of the *Wolbachia* hosts (Figure 5D). Moreover, it is also gene richer than the two other published Collembolan genomes of *Holacanthella duospinosa* where they found 12,000 gene models (Wu *et al.*, 2017) and *Orchesella cincta* with 20,249 genes (Faddeeva-Vakhrusheva *et al.*, 2016). In addition, groups of genes related to DNA repair expanded in both genomes, with *F. candida* having expanded groups of ATP-dependent DNA helicases which are important for DNA repair (Wu and Hickson, 2006; Faddeeva-Vakhrusheva *et al.*, 2017). Finally, both host and endosymbiont contain high amounts of foreign DNA, as 2.8% of *F. candida* genes are horizontally transferred genes (HTG) from a wide variety of organisms including bacteria, fungi and protists, but not from *Wolbachia*. This might sound low compared to the 25.4% of phage DNA in *w*Fol but is in fact the highest percentage found in any metazoan genome except for rotifers and some nematode species (Faddeeva-Vakhrusheva *et al.*, 2017) Moreover, this abundance of mostly intact HTG was correlated with an increase in transposons. Thus, also the *F. candida* genome seems focused on preserving genes. This is as far as we know, the first case that shows signs of convergent evolution between endosymbiont and host genomes. In addition, the correlation between *Wolbachia* gene number and host gene number of all combinations with usable genomes, suggest that convergence between host and endosymbiont genomes may have taken place in more *Wolbachia* strains.

Regions of phage origin

The integration of phage DNA is one of the main reasons for the large size of the *w*Fol genome. The RPO included three phage WO regions and two phage WO-like islands, which is within the range of two to five prophage WO haplotypes found in

other fully sequenced Wolbachia genomes (Kent *et al.*, 2011). However, the size of the *w*Fol WO regions is much larger and two of the regions contain multiple copies of essential phage gene clusters. Possibly, multiple phages clumped together within the *w*Fol genome or the phages that integrated had multiple copies of the same modules. Alternatively, a recent duplication took place, but this is not very likely because the collinear blocks between phages were very fragmented and the longer blocks contained mainly the conserved modules found in all WO phages. Such large clusters containing multiple sets of the same phage modules have previously been found in *Wolbachia* genomes *w*Pip and *w*No (Klasson *et al.*, 2008; Ellegaard *et al.*, 2013) and hamper an assessment of the exact number of phages that integrated within this genome.

All three WO regions within *w*Fol contained the characteristic elements and standard modules of phage WO. These included the Patatin gene that is probably helpful for entering the host cell (Kent *et al.*, 2011) and the recently defined eukaryotic association module including latrotoxin-CTD domain proteins, which are related to black widow venom genes (Bordenstein and Bordenstein, 2016). Neither of the WO regions contained the *cifA* and *cifB* genes linked to CI (Le Page *et al.*, 2017), therefore it is very unlikely that *w*Fol can cause CI. Some *w*Fol RPO features are exceptional, such as the vast number of resolvases and transposases, which are thought to be involved with phage integration into their bacterial host (Campbell, 1992; Groth and Calos, 2004). However, integration by transposases usually takes place at non-specific integration sites while resolvases use conserved sites (Campbell, 1992). Thus, the excessive presence of both transposases and resolvases in the *w*Fol phages raises questions concerning their mode of integration. Finally, a toxin-antitoxin system that is new for *Wolbachia* was found in the WOFol2 and 3. This AbiEi and AbiEii system is an abortive infection system that can cause altruistic cell death to prevent new phages from settling within the bacterial population (Dy *et al.*, 2014). One might argue that such a system may prevent further genome expansions by additional phages.

Genetic basis of parthenogenesis induction

We also perused the *w*Fol genome for genes that are associated with PI to elucidate the genetic basis of this reproductive manipulation. Together with three other PI *Wolbachia*, *w*Uni of supergroup A and *w*Tpre and *w*Lcla of supergroup B, they were compared to a diverse set of 14 *Wolbachia* including members of 6 different supergroups (A,B,C,D,F and L). We searched for gene sets present in at least three PI and a maximum of two other *Wolbachia* strains, comparable to the CI phenotype where a toxin-antitoxin gene cluster was shared by all CI inducing or repressing strains and some others (Beckmann, Ronau and Hochstrasser, 2017; Le Page *et al.*, 2017). We did not find any genes present in all PI strains, which may have several non-mutually exclusive explanations. First, PI genes may not be uniquely present in

Wolbachia expressing the PI phenotype, but also in some of the *Wolbachia* causing other reproductive manipulations, similar to what is seen for the CI phenotype (Lindsey *et al.*, 2018). The two genes we identified to be associated with three of the PI-strains were also present in *w*Bolb-1 and in *w*Pip_Pel that are both in supergroup B. The presence of these genes in the two other strains might have to do with the fact that phenotypic expression of *Wolbachia* has been found to be determined in combination with the host genotype. For example, the CI-strain *w*Rec caused male killing after being transferred to a new host (Jaenike, 2007) and *w*Tei caused CI after a host transfer while showing no signs of reproductive manipulation in its natural host (Zabalou *et al.*, 2008). This same interaction with host genotype may apply to PI expression. A second possibility is that the PI phenotype is not a single genotype but is achieved via different cellular mechanisms formed by convergent evolution, in which case we do not expect a single gene cluster to be shared by all PI-*Wolbachia*. Support for this explanation can be found in studies showing that parthenogenesis is induced in a two-step mechanism of diploidization of the gamete followed by feminization of the egg (Tulgetske, 2010; Ma *et al.*, 2014). Diploidization can be achieved in different ways. In both *Trichogramma* wasps and *Leptopilina clavipes* (the hosts of respectively *w*Tpre and *w*Lcla, the two PI inducing *Wolbachia* of supergroup B) gametes are rendered diploid by failure of the first mitotic anaphase division (Stouthamer and Kazmer, 1994; Pannebakker *et al.*, 2004). In *Muscidifurax uniraptor* (host of *w*Uni of supergroup A) diploidization is only restored after the second mitotic division through a fusion of the two mitotic nuclei (Gottlieb *et al.*, 2002). For, *w*Fol of supergroup E the speculated mechanism is diploidy restoration through either non-disjunction or fusion of the two haploid division products at the end of the first anaphase (Koivisto and Braig, 2003) and thus similar to the situation in *w*Tpre and *w*Lcla, although the difference in sex determination system might call for a different manipulation for the same outcome. Therefore, it has already been suggested that diploidy restoration in *Wolbachia* most likely has a polyphyletic origin (Pannebakker *et al.*, 2004). With this in mind, the cluster of two genes uniquely shared by *w*Tpre and *w*Lcla might still be very interesting and might be linked to one of the two steps of PI in these lines that seem to have a similar mechanism developed in supergroup B.

Conclusions

Large RPOs with ample repair genes and accumulation of repetitive and transposable elements make up most of the expansion of the *w*Fol genome. This genomic signature of gene conservation was mirrored in the *F. candida* host genome. We found that a large part of the variation within the genome size and gene number of facultative endosymbionts is correlated to the gene number of the host. This suggests that host and symbiont genome may be subject to correlated

selection pressures that resulted in convergent evolution between host and endosymbiont, or that somehow the host may have a direct influence on the symbiont genomes. However, these selection pressures would probably be neutralized after endosymbionts move into a specialized cell, explaining the well-documented steady genome decrease seen in those cases (Moran and Mira, 2001; Bennett and Moran, 2013; Moran and Bennett, 2014). Therefore, genome reduction may ultimately result from becoming obligate. However, before the endosymbiont resides in a specialized cell, but while being mainly transmitted vertically, genome expansion of endosymbiont genomes might occur.

The search for the PI genes yielded a set of potential candidate genes. Elaborating on these findings could answer whether the PI genes are monophyletic or polyphyletic. The complexity of this trait and the indications that this is caused by multiple genes could also means that both are not mutually exclusive in this case.

Methods

Assembling of wFol and wLcla

Sequencing and assembling of the *w*Fol genome was described in Faddeeva-Vakhrusheva et al. (2017). We corrected this assembly using the Illumina HiSeq 2000 data of Gerth et al. (2014) (NCBI accession number: SRR1222159). Illumina reads were mapped to the assembly with BWA (Li and Durbin, 2009) using default parameters. Variants between the assembly and the mapped Illumina reads were called and quality filtered using GATK v. 3.7 (Depristo *et al.*, 2011), filtered based on read depth with vcffilter ("DP > 10") of the Vcflib package (E. Garrison, 2012, https://github.com/vcflib/vcflib). Variants were inspected manually and when they were supported by the mapped Illumina data they were included in the genomic sequence.

Reads used for the *w*Lcla assembly were taken from a study that sequenced its host *Leptopilina clavipes* (Kraaijeveld *et al.*, 2016) using both an Illumina HiSeq (HiSeq) and a Pacific Biosciences RS I (PacBio) library. Hiseq reads were used to error correct the PacBio reads with the PacBioToCA pipeline of CeleraAssembler7.0 (Myers *et al.*, 2000). We used the corrected PacBio reads for the *w*Lcla assembly. To filter out the *Wolbachia* reads, all PacBio reads were mapped onto the *w*Tpre and the *w*Pip_Pel genome with BLASR (Chaisson and Tesler, 2012). A consensus based on the mapped reads was made per reference genome with PBDAG-Con (Chin *et al.*, 2013) and the resulting contigs were extended and joined with PBjelly (English *et al.*, 2012). Thereafter, all reads were mapped back to these preliminary assemblies and PBDAG-Con and PBjelly were run again to extend the preliminary contigs further. All reads mapping to these two assemblies were extracted and assembled de novo in Mira (Chevreux, Wetter and Suhai, 1999) with the following

parameters: COMMON_SETTINGS –SK:mmhr=1 PCBIOHQ_SETTINGS –CO:mrpg=5. Next, all corrected Pacbio reads were mapped back to the Mira assembly, after which two iterations of the combination of PBDAG-Con and PBjelly were run to extend and connect the contigs as much as possible. The process ended with a final step of PBDAG-Con to confirm the extensions based on mapped reads. Finally, the HiSeq data was mapped to the assembly, variants were called, manually checked and the assembly was adjusted in the same way as described for the *w*Fol assembly.

Annotation of wFol and wLcla

The annotation of both genomes was done using the DIYA pipeline (Stewart, Osborne and Read, 2009), in which we included: Prodigal for gene prediction (Hyatt *et al.*, 2010), tRNAscan-SE and RNAmmer to predict RNA features (Lowe and Eddy, 1997; Lagesen *et al.*, 2007) and GenePRIMP to mark possible pseudogenes and short genes without annotation (Pati *et al.*, 2010). Predicted genes smaller than 100 amino acids without hits in any of the databases were removed. Genes were manually annotated as pseudogenes when they contained frameshift mutations, premature stop-codons, missing start codons or disruptions by IS-insertions. In addition, protein domains were predicted according to the Pfam database (Finn *et al.*, 2014), BlastP to the NCBI database (Altschul *et al.*, 1990) and FASTA searches (Pearson and Lipman, 1988) against a well-curated in-house database by Lisa Klasson of *Wolbachia* genomes. All results were loaded into Artemis (Rutherford *et al.*, 2000) in which they were manually curated. Transposases were blasted to the IS-finder database to determine the family and group (https://www-is.biotoul.fr). Uncharacterized/unique proteins were run through the InterPro databases with InterProScan (Jones *et al.*, 2014). Genes that did not exhibit features indicative of any specific function, but did contain transmembrane segments, cytoplasmic and non-cytoplasmic domains were characterised as putative membrane proteins.

For *w*Fol, equal distribution of gene groups and transposase families over chromosomal parts (74.6%) and RPOs (25.4%) was tested with binomial tests in R (R Core Team, 2017).

Annotation completeness and ortholog identification

We compared the *w*Lcla and *w*Fol assemblies to a set of 16 other *Wolbachia* genomes, selected based on assembly quality as well as phylogenetic and functional diversity (Supplementary table 3)(Wu *et al.*, 2004; Foster *et al.*, 2005; Klasson *et al.*, 2008, 2009; Darby *et al.*, 2012; Duplouy *et al.*, 2013; Ellegaard *et al.*, 2013; Metcalf *et al.*, 2014; Nikoh *et al.*, 2014; Sutton *et al.*, 2014; Brown *et al.*, 2016; Cotton *et al.*, 2016; Leclercq *et al.*, 2016; Lindsey *et al.*, 2016; Newton *et al.*, 2016). Protein sequences of these 16 genomes were downloaded from the NCBI database and the completeness of their gene content was predicted with the BUSCO v3

pipeline, which compares the *Wolbachia* genomes to a set of 148 single copy bacterial genes (Bacteria odb9)(Simão *et al.*, 2015). Furthermore, orthogroups were inferred using OrthoFinder and are defined as groups of genes that all derived from a single gene in the last common ancestor (Emms and Kelly, 2015). This allowed us to look for orthogroups shared between *Wolbachia* associated with PI and find unique genes (genes lacking orthologues in other *Wolbachia*).

In all genomes, expanded and contracted orthogroups were identified by calculating the z-scores (Cao *et al.*, 2013), which is calculated by subtracting the average number of genes in an orthogroup over all species from the gene number of a focal species and dividing this by the standard deviation. A z-score of 2 or above indicates an expansion and a score of -2 or below a contraction of an orthogroup.

Prophage annotation and synteny

Prophage regions of *w*Fol and *w*Lcla were initially identified by PHASTER (Arndt *et al.*, 2016). While this correctly finds the more common phage genes it does not identify the more diverse EAM (Bordenstein and Bordenstein, 2016), as they contain of eukaryotic genes picked up by phages that are therefore not being recognised as phage genes. Therefore, the EAM were defined during manual curating of the prophage regions. Phage region lengths of other phages were based on annotations from Bordenstein & Bordenstein (2016) to include the whole phages with EAM or, if not available from that paper, by determining them in a similar way (Supplementary table 1). As, the phage region are often the most difficult parts to assemble, this was only done for *Wolbachia* assemblies with two or less scaffolds, to avoid the use of incomplete sets of RPOs within our analysis. Moreover, *w*Cle could not be used as phage genes seemed to be removed from its annotation.

Collinearity between the regions of phage origin (RPOs) was analysed with the MCScanX package (Wang *et al.*, 2012). A BlastP search of all against all RPO protein sequences was performed with an E-value cutoff of 1e−10, followed by a MCScanX run (-m 2). Synteny plots were drawn using Circos (Krzywinski, 2009).

Whole genome correlations

To examine overall genomic expansion and contraction patterns and their possible connection to phages and their host, we assessed correlations between *Wolbachia* genome size versus amount of phage DNA, *Wolbachia* genome size versus host genome size, and number of *Wolbachia* genes versus number of host genes for all combinations that had sufficient reliable data. Genomic information of all available hosts of the *Wolbachia* analysed in this study was collected from online servers, most were downloaded from NCBI (hosts and accession numbers: *Brugia malayi,* GCA_000002995.4; *Cimex lectularius,* GCF_000648675.1; *Culex quinquefasciatus,*

GCA_000209185.1; *Drosophila melanogaster,* GCF_000001215.4; *Drosophila simulans,* GCA_000754195.2; *Folsomia candida,* GCA_002217175.1; *Onchocerca ochengi,* GCA_000950515.2 and *Trichogramma pretiosum*, GCF_000599845.1) and data for *Onchocerca volvulus* strain Cameroon was collected from http://parasite.wormbase.org/Onchocerca_volvulus_prjeb513/Info/Index/. Spearman's rank correlations were calculated in R.

List of abbreviations

PI = parthenogenesis induction
CI = cytoplasmic incompatibility
CDS = protein coding sequences
RPO = region of phage origin
ANK= ankyrin repeat containing protein
EAM = eukaryotic association module

Declarations

Ethics approval and consent to participate
Not applicable

Availability of data and material

Assemblies of *w*Fol and *w*Lcla are available on GenBank under respectively accession number GCA_001931755.2 and GCA_001855655.1 and including annotation on the data portal of Vrije Universiteit Amsterdam http://www.collembolomics.nl. In addition, the wFol and wLcla annotations are also added to the additional files (see Additional file 7 and 8). Assemblies of other *Wolbachia* genomes used for comparisons can be found on GenBank, accession numbers are provided in Additional file 2: Supplementary table 2.

Funding

This work is supported by the Netherlands Organisation for Scientific Research, VICI grant 865.12.003. DR received additional funding from NWO-TTW grant number 15494. These funding bodies had no role in the design of the study and collection, analysis, and interpretation of data or in writing this manuscript.

Authors' contributions

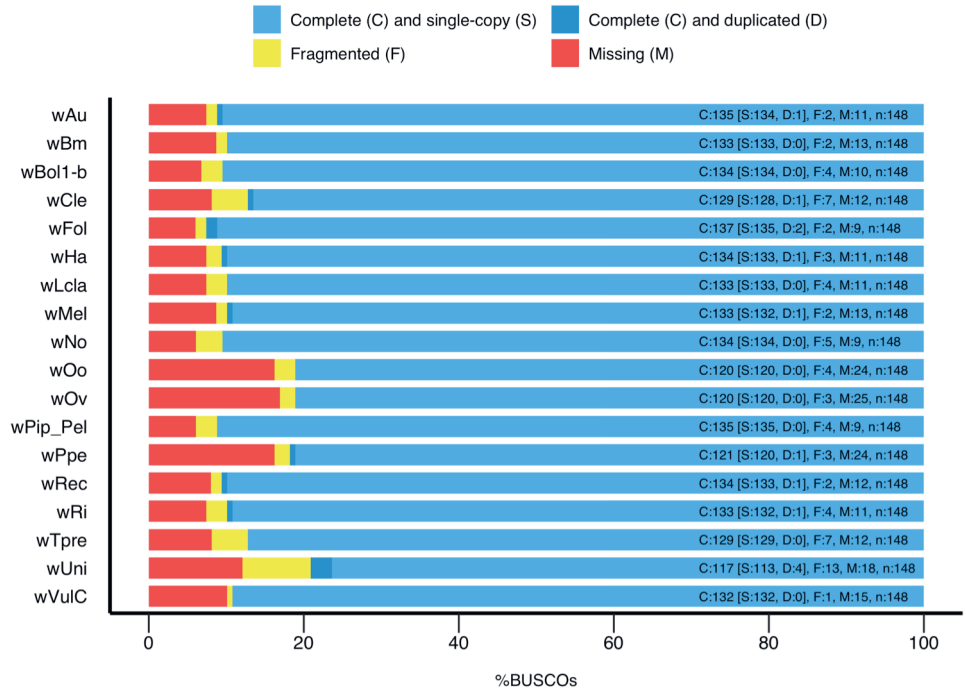
AAK, LK, KK and JE conceived, designed, coordinated the study and drafted the manuscript. AAK, KK, SYA and RHAMV performed the assembly of the genomes. AAK and LK annotated the genomes. AAK and DR performed orthology analysis and gene family expansion analysis. AAK and KK analysed phage synteny. AAK calculated and created the correlation plots. AAK and JE searched for genes related to PI. All authors provided input and comments on subsequent revisions of the manuscript and approved the final version. Finally, each of the authors has

confirmed to be accountable for all aspects of the work by appropriately investigating and resolving issues related to accuracy and integrity.

Acknowledgements
We want to thank Peter Neleman for bioinformatical support, SURFsara for the use of their HPC cloud and Sarah and Seth Bordenstein for their advice on phage annotation.

Additional files

Additional file 1
Title: BUSCO assessment results of 18 *Wolbachia* genomes included in this paper
Description: Protein sequences of the 18 *Wolbachia* strains were searched for a set of 148 single copy bacterial genes (Bacteria odb9), defining the complete, fragmented and missing genes.



Additional file 2
Title: Supplementary tables on *Wolbachia* strains
Description: Supplementary table 1 on RPO regions within the *Wolbachia* strains analysed and Supplementary table 2 with extra information on these strains.

Additional file 4
Title: Results OrthoFinder analysis
Description: This file contains multiple sheets with selected orthogroups derived from the Orthofinder analysis as well as the z-scores. (XLSX 1.3Mb)
Link: https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-019-5492-9/MediaObjects/12864_2019_5492_MOESM4_ESM.xlsx

Supplementary table 1: Characteristics of RPOs within the *Wolbachia* strains used in this study with fewer than three scaffolds and properly annotated phage genes.

Wolbachia	Phage	Locus-tags	Length (bp)	Total (bp)
*w*Au	WOAu1-2	WPAU_0243-WPAU_0329	67.918	121.654
	WOAu3	WPAU_0623-WPAU_0672	53.736	
*w*Bm strain TRS	NA	NA	NA	NA
*w*Fol	WOFol1	wFol_02640-wFol_03410	83.672	458.452
	WOFol2	wFol_05200-wFol_06400	132.270	
	WOFol3	wFol_08880-wFol_10640	215.744	
	WO-Island 1	wFol_03590-wFol_03680	23.233	
	WO-Island 2	wFol_06910-wFol_06950	3.533	
*w*Ha	WOHa1	wHa_02010-wHa_02760	100.443	142.542
	WOHa2	wHa_03390-wHa_03840	42.099	
*w*Mel	WOMelA	WD0253-WD0296	38.175	144.430
	WOMelB	WD0563-WD0647	87.451	
	WO-Island	WD0507-WD0514	18.804	
*w*No	WONo1	wNo_01000-wNo_01400	41.397	161.733
	WONo2	wNo_07170-wNo_07380	13.585	
	WONo3	wNo_09000-wNo_09150	21.354	
	WONo4	wNo_10070-wNo_10350	47.238	
	WO-Island 1	wNo_01980-wNo_02200	31.868	
	WO-Island 2	wNo_05070-wNo_05150	6.291	
*w*Oo	NA	NA	NA	NA
*w*Ov	NA	NA	NA	NA
*w*Pip_Pel	WOPip1-3	WP0236-WP0354	129.469	277.399
	WOPip4	WP0405-WP0465	71.818	
	WOPip5	WP1289-WP1352	76.112	
*w*Ri	WORiA	wRi_012450-wRi_012680	22.148	248.301
	WORiB	wRi_005310-wRi_005930	73.575	
	WORiB(2)	wRi_009980-wRi_010580	72.115	
	WORiC	wRi_006620-wRi_007230;	67.515	
		wRi_007550-wRi_007660	12.948	
*w*Tpre	WO-Island 1	wTpre_69-76	3.163	28.317
	WO-Island 2	wTpre_1393-1428	25.154	

Supplementary table 2: Characteristics of *Wolbachia* strains used for comparisons.

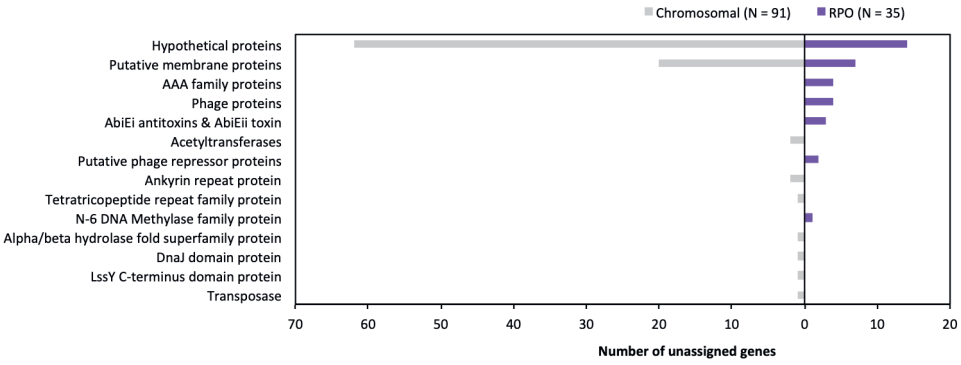
Strain	Sg	Host	Effect on host	f/o	Reference	Accession nr.
*w*Au	A	*Drosophila simulans*	viral protection	f	Sutton et al. 2014	GCA_000953315.1
*w*Ha	A	*Drosophila simulans*	CI	f	Ellegaard et al. 2013	GCA_000376605.1
*w*Mel	A	*Drosophila melanogaster*	low CI/ antiviral	f	Wu et al. 2004	GCA_000008025.1
*w*Rec	A	*Drosophila recens*	CI/MK other host	f	Metcalf et al. 2014	GCA_000742435.1
*w*Ri	A	*Drosophila simulans*	CI	f	Klasson et al. 2009	GCA_000022285.1
*w*Uni	A	*Muscidifurax uniraptor*	PI	o	Newton et al. 2016	GCA_001983635.1
*w*Bol1	B	*Hypolimnas bolina*	MK/CI res	f	Duplouy et al. 2013	GFA_000333775.1
*w*Lcla	B	*Leptopilina clavipes*	PI	o	this study	GCA_001855655.1
*w*No	B	*Drosophila simulans*	CI	f	Ellegaard et al. 2013	GCA_000376585.1
*w*Pip	B	*Culex quinquefasciatus*	CI	f	Klasson et al. 2008	GCA_000073005.1
*w*Tpre	B	*Trichogramma pretiosum*	PI	o	Lindsey et al. 2016	GCA_001439985.1
*w*VulC	B	*Armadillidium vulgare*	F	f	Leclercq et al. 2016	GFA_001027565.1
*w*Oo	C	*Onchocerca ochengi*	normal development	o	Darby et al. 2012	GCA_000306885.1
*w*Ov	C	*Onchocerca volvulus*	normal development	o	Cotton et al. 2016	GFA_000530755.1
*w*Bm	D	*Brugia malayi*	normal development	o	Foster et al. 2005	GCA_000008385.1
*w*Fol	E	*Folsomia candida*	PI/egg hatching	o	this study	GCA_001931755.2
*w*Cle	F	*Cimex lectularius*	vitamin B production	o	Nikoh et al. 2014	GCA_000829315.1
*w*Ppe	L	*Pratylenchus penetrans*	NA	f	Brown et al. 2016	GCA_001752665.1

Abbreviations are used for certain host manipulations: CI for cytoplasmic incompatibility, PI for parthenogenesis induction, F for feminization and MK for male killing. Column label abbreviations: Sg stands for supergroup, f/o stands for facultative or obligate. If information was not available these characteristics were labelled NA.

Additional file 5

Title: Distribution of unassigned *w*Fol genes

Description: Double sided histogram that presents the distribution of unique *w*Fol genes between the chromosomal and the regions of phage origin (RPO).

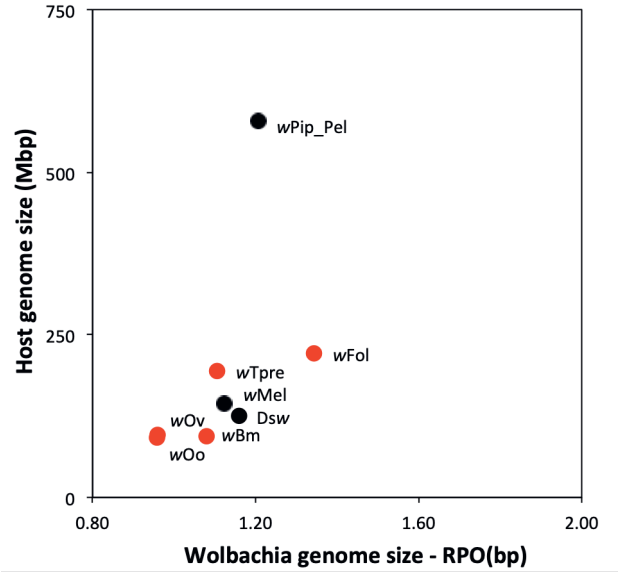


Additional file 3

Title: Abbreviations Figure 2 Description: List of abbreviations of annotations within arrows in Figure 2.

AAA	AAA family protein
ABCt	ABC transporter permease/ATP-binding protein
ado	L-allo-threonine aldolase
amt	adenosine monophosphate-protein transferase
ANK	ankyrin repeat containing protein
ANK+	ankyrin repeat containing protein with additional domains
dhg	UDP-glucose 6-dehydrogenase
Dnar	putative DNA repair protein
end	putative endonuclease/ HNH endonuclease family protein
GpA	Phage terminase large subunit GpA
GpC	Putative phage minor capsid protein C
GpD	Putative phage head decoration protein D
GpE	Putative phage major capsid protein E
GpJ	Putative phage baseplate assembly protein J
GpU	Phage tail assembly protein GpU
GpV	Phage baseplate assembly protein V
GpW	Putative phage baseplate assembly protein GpW
GpX	Phage tail protein X
GpZ	Putative phage minor tail protein Z
Heli	Helicase SNF2 family
hsW	GpW-like phage protein
LCD	Phage late control gene D protein
ligA	DNA ligase (NAD(+)) LigA
met	Phage related DNA methylase
mfft	major facilitator family transporter
mtf	methyltransferase
PAAR	PAAR-family protein
pap	phage anti-repressor protein
Pat	patatin
phc	putative phage head-tail connector protein
phyH	Phytonoyl-CoA dioxygenase PhyH
pld	phospholipase D-like protein
plp	Putative phage portal protein, lambda family
pmp	putative membrane protein
ppp	putative phage related protein
ppr	putative phage repressor
ppt	putative phage tail protein
pseu	pseudogene
pss	putative type II secretion system protein
ptp	Putative phage tail protein
radC	DNA repair protein radC
res	putative resolvase
Res	resolvase
rhuM	putative virulence factor RhuM
rnap	putative RNA polymerase
rnh	ribonuclease H-like protein
rSAM	radical SAM protein
S-70	sigma 70
TAC	Phage tail assembly chaperone protein
tox	toxin
tra	transposase
trf	nucleotide-diphospho-sugar transferase
TSP	Phage major tail sheath protein
TTM	Phage tail tape measure protein
TTP	Phage tail tube protein
ucp	uncharacterized protein
uni	unique protein
vir	putative virulence protein
WDr	WD repeat domain protein

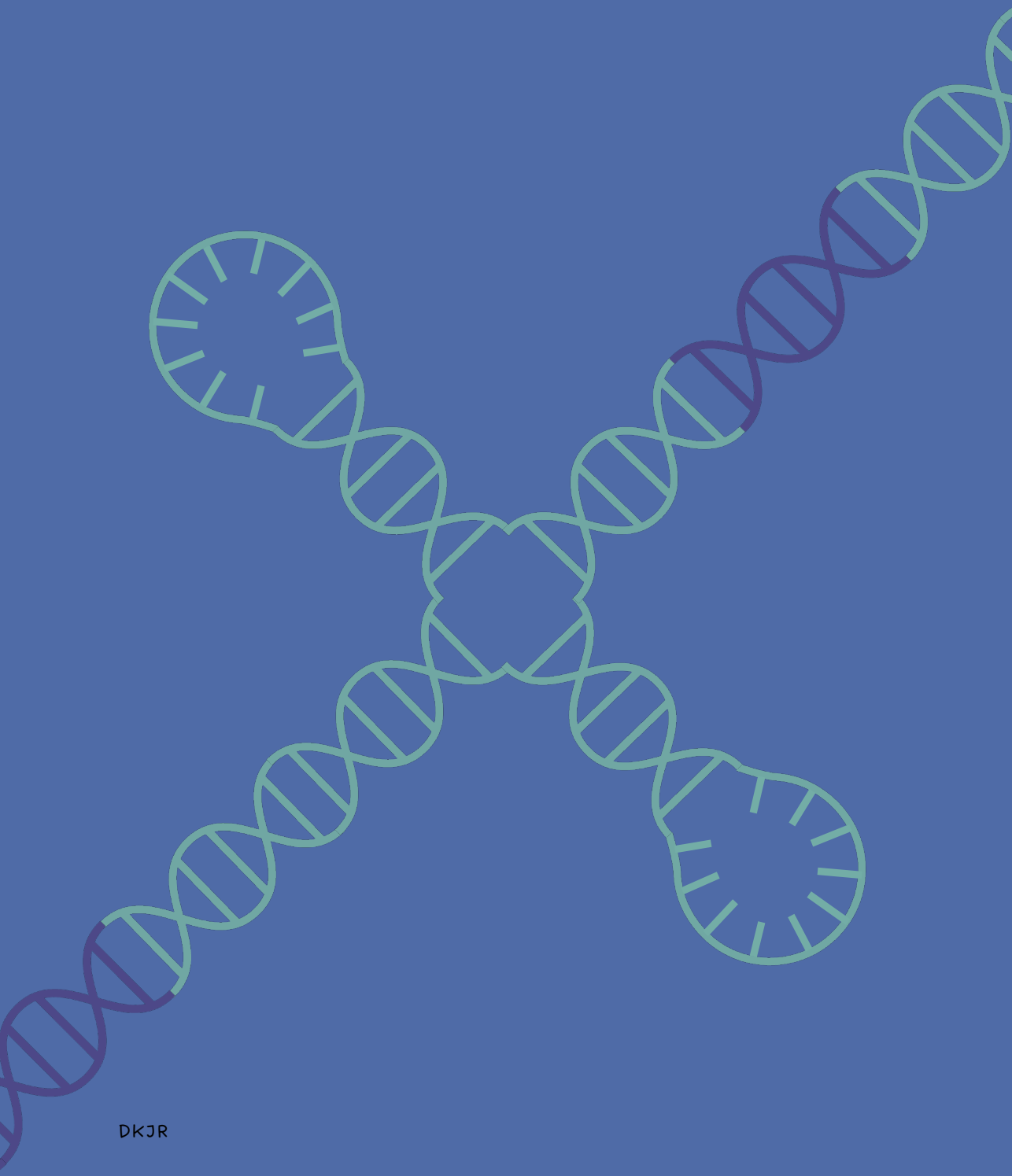
Additional file 6
Title: Host genome size to *Wolbachia* genome size excluding RPOs
Description: Correlation between host genome size and *Wolbachia* genome size excluding RPOs. Red data points indicate obligate transitional and black facultative endosymbionts; all points are labelled with strain names and for the correlations with the host the four strains residing in *Drosophila simulans* were averaged and labelled as Ds*w*.



Additional file 7
Title: *w*Fol annotation
Description: GenBank data file with the *w*Fol annotation described in this paper. (GBK 3Mb)
Link: https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-019-5492-9/MediaObjects/12864_2019_5492_MOESM7_ESM.gbk

Additional file 8
Title: *w*Lcla annotation
Description: GenBank data file with the *w*Lcla annotation described in this paper. (GBK 2.3Mb)
Link: https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-019-5492-9/MediaObjects/12864_2019_5492_MOESM8_ESM.gbk

CHAPTER 5



DKJR

Contrasting genome dynamics in sexual and asexual lineages of F. candida

A.A. Kampfraath, E. Jonkheer, S. Smit, S.Y. Anvar, J. Mariën, J. Ellers and K. Kraaijeveld

Status: unpublished

Abstract: The transition from sexual to asexual reproduction is a significant evolutionary event that has been observed in various organisms. However, the underlying mechanisms and genomic changes associated with this transition are not yet fully understood. Palindromic regions in the first genome of an asexual lineage of *Folsomia candida* might have been the consequence of this transition. To answer this question and to better understand the effects of a transition to asexual reproduction on a genome, we investigated the differences between sexual and asexual lineages of *F. candida* using comparative genomics.

We sequenced, assembled and annotated the genomes and mitochondrial genomes of two asexual and one sexual lineage of *F. candida* and screened them for repetitive elements (REs) and large segmental duplications, including palindromic regions. Thereafter, we determined whether genome dynamics differed between the sexual and asexual lineages by looking into the changes of these elements over time. Furthermore, we assessed the presence of palindromes in all three lineages.

Our comparative genomic analysis revealed that both sexual and asexual lineages possess palindromic regions, indicating this is not a sign of asexuality in *F. candida*. However, profound differences in genome architecture between the sexual and asexual lineages of *F. candida* were revealed. While the sexual lineage possesses a smaller and more streamlined genome, the asexual lineages exhibit signs of genetic expansion, characterized by significantly elevated levels of repetitive elements (REs), segmental duplications (SDs), and mitochondrial DNA remnants (NUMTs). This pattern suggests that transitioning to an asexual mode of reproduction is associated with a unique genomic strategy: one that promotes both accumulation and retention of diverse genetic material. Specifically, the ability of asexual lineages to tolerate and retain large amounts of repetitive DNA (REs, SDs) and integrate mitochondrial material (NUMTs) may represent an adaptive evolutionary strategy to overcome need of sexual recombination.

Key words: *Folsomia candida*, transposable elements, palindromes, segmental duplications, NUMTs, Arthropoda, Hexapoda, Collembola

Introduction

With the term genome dynamics, the changes in genome structure and content over time are described. This includes the acquisition, rearrangement, and deletion of genetic material. Moreover, it recognizes genomes as inherently dynamic rather than static entities that respond to evolutionary forces and ecological pressures (Parfrey, Lahr and Katz, 2008; Ochman and Davalos, 2006). Genome dynamics are not constant over time but fluctuate in response to evolutionary pressures, environmental shifts developmental stages or shifts in reproductive strategy, as evidenced by episodic bursts of genomic instability across generations (Barrick and Lenski, 2013; Hollister *et al.*, 2015; Heasley and Argueso, 2022; Vargas-Chávez *et al.*, 2025).

Fluctuations in genome dynamics can be caused by changes in the rate or the source of new genetic material, and the extent to which such material is preserved in the genome. For instance, a periodic burst of transposon activity or a whole genome duplication may cause an increase in genome size (Zwaenepoel and Van de Peer, 2019; Roelofs *et al.*, 2020). These transposons are part of the repetitive elements (REs) that can take up large proportions of the eukaryotic genome and play a key role in insect genome evolution (Gilbert, Peccoud and Cordaux, 2021). A recent large-scale analysis of over 600 insect genomes revealed rapid coevolution between repetitive elements and protein-coding genes at their genomic interfaces, demonstrating dynamic interactions that shape insect genome architecture (Sproul *et al.*, 2023).

Fluctuations in genome dynamics are often associated with life-style transitions or speciation events, but currently it is not clear if genomic rearrangements are a cause or consequence of these major evolutionary changes (Faria and Navarro, 2010; Siozios *et al.*, 2024). The transition to asexual reproduction is a major lifestyle change that is commonly believed to have major genomic consequences (Jaron *et al.*, 2020). Asexuality has frequently been reported to affect gene content and diversity, through changes in heterozygosity, mutation accumulation, horizontal gene transfer and gene family expansion (Lovell *et al.*, 2017; Jaron *et al.*, 2020). However, it has been difficult to draw general conclusions about the genetic consequences of transitioning to asexual reproduction, as the different studies show inconsistent changes in the presence or absence of specific genetic elements mentioned above (Tvedte, Logsdon and Forbes, 2019; Jaron *et al.*, 2020). A focus on the larger scale genome dynamics, like structural variants and genome duplications, instead of specific genetic elements might give us better insight into the genomic changes in asexual species.

Major improvements in the availability and affordability of long-read sequencing techniques have produced highly contiguous genome assemblies that enable shifting the focus in evolutionary genomics from gene content to genome structure and dynamics. The high contiguousness of these assemblies can result in

a much better detection of structural changes like increases of REs, with increased detection of up to 162% of REs in long read assemblies (Sproul *et al.*, 2023).

In addition to REs, studies on genome dynamics have identified partial or whole genome duplications (De Bodt, Maere and Van de Peer, 2005; Van de Peer, Maere and Meyer, 2009; Vanneste, Van de Peer and Maere, 2013; Zwaenepoel and Van de Peer, 2019), inter and intra species variation in genome synteny (Simakov *et al.*, 2022; Francis, Vargas and Wörheide, 2025), and deletions (Mesbah-Uddin *et al.*, 2018; Jonkheer *et al.*, 2025), demonstrating the array of large-scale genomic elements that can be influenced by genome dynamics.

Multiple studies have reported the occurrence of segmental duplications within genomes of asexual organisms (Flot *et al.*, 2013; Blanc-Mathieu *et al.*, 2017; Faddeeva-Vakhrusheva *et al.*, 2017; Nowell *et al.*, 2018). Strikingly, several of these studies found palindromes, a special kind of segmental duplication, in the genomes of asexual organisms (rotifers, nematodes and springtails). Palindromes are duplicated regions on a single chromosome, that can align and form hairpins because of their reverse orientation. This mechanism allows for gene conversion within duplicated regions and has been shown to purge mutations from Y-chromosomes in humans and chimpanzees (Rozen *et al.*, 2003; Hughes *et al.*, 2005). The formation of hairpins could also hinder chromosome pairing during meiosis, which is the reason why palindromic regions are thought to be an indication of areas of the genome that do not recombine, like Y-chromosomes and chromosomes of ameiotic lineages. However, palindromes do not seem to be consistently present in all asexual genomes that have been studied (Jaron *et al.*, 2020), although part of this result may depend on the quality and contiguousness of the genome assemblies that determine the ability to detect palindromes. Therefore, the relationship between the presence of palindromic regions and asexuality is still unresolved, and leaves open the possibility that the presence of palindromes is a species-specific trait rather than generally associated with asexuality.

One of the asexual species with a substantial number of palindromic regions in its genome is the springtail *Folsomia candida* (Faddeeva-Vakhrusheva *et al.*, 2017; Jaron *et al.*, 2020). Eleven to fifteen palindromes were detected in the sequenced asexual lineage of *F. candida* which co-occur with transposable elements (TE) and horizontally transferred genes, indicating that they are in regions with elevated genomic dynamics. For this species both asexual and sexual linages have been described (Frati *et al.*, 2004; Fountain and Hopkin, 2005; Krogh *et al.*, 2008). Comparing the genomes of sexual and asexual *F. candida* linages could provide us with indications whether palindromes within non-sex chromosomes are interrelated to asexuality.

Hence, for this study we sequenced, assembled and annotated the genomes and mitochondrial genomes of two asexual and one sexual lineage of *F. candida* and screened them for REs and large segmental duplications, including

palindromic regions. Thereafter, we determined whether genome dynamics differed between the sexual and asexual lineages by looking into the changes of these elements over time. Furthermore, we assessed the presence of palindromes in all three lineages to investigate if palindromes are linked to asexuality or if they are species-specific.

Methods

Folsomia cultures

The two parthenogenetic and single sexual lineages of *F. candida* used for this study were previously described in Kampfraath et al., 2020.The parthenogenetic BE strain ("Berlin strain", Vrije Universiteit, Amsterdam) is a well-characterized asexual strain that is commonly used for lab tests. The second parthenogenetic lineage used in this study was collected in a ground water sample in October 2013 in Evergem (EV), Belgium, and since kept in the lab of the Vrije Universiteit Amsterdam. The sexual population CR was collected in Cremona (Italy) and described by (Frati *et al.*, 2004). Before DNA isolation, this strain was inbred by crossing one male and one female and thereafter keeping only the first egg clutch of the next 12 generations to ensure sib-sib matings. This was done to minimize the amount of genetic variation within this population to allow for a better genome assembly. All cultures were maintained in climate rooms under stable conditions in pots with a water-saturated plaster of Paris bottom containing 10% charcoal and were fed dried baker's yeast (Dr. Oetker) *ad libitum*.

Sample preparation and sequencing

As the genome of the BE lineage was already sequenced, high molecular weight DNA (HM-DNA) was only extracted for the EV and CR lineages. For the HM-DNA isolation, eight times 100 animals per lineage were crushed in CTAB lysis buffer containing 0,2 % β-mercaptoethanol (800 µl buffer per 100 animals) and lysates were incubated for 2 hour at 60°C. HM-DNA was extracted from the lysates with phenol:chloroform:isoamyl alcohol (25:24:1) and two times chloroform:isoamyl alcohol (24:1). The water phase samples were pooled into respectively 4 and 2 samples per lineage, and HM-DNA was precipitated with isopropanol at 4 °C. Pellets were dried, dissolved in 400 µl H_2O and incubated for 15 minutes at 37°C with 1 µl RNAse (100mg/ml). HM-DNA was purified again with chloroform:isoamyl alcohol (24:1) and precipitated with (3 x volume) ethanol 100% and (1/10 x volume) NaAc (3M, pH=5.2). Pellets were washed with ethanol 70%, dried and dissolved in 200 µl H_2O. A final cleanup of the HM-DNA was necessary for downstream procedures. For this we used the MoBio Power Clean® DNA clean up kit and followed the manufacturer's protocol. Fragment size of the HM-DNA was checked on a 1 %

agarose gel and concentrations of DNA and RNA were measured using a Qubit fluorometer (Thermo Fisher Scientific). For the genome of the EV lineages, fragments of more than 20 kbp were sequenced using 12 SMRT cells on the PacBio RS II platform (with P6 chemistry) and the CR lineage was sequenced on the PacBio Sequel platform using 10 SMRT cells both according to the manufacturer's protocol at the Leiden Genome Technology Center (LGTC).

Assembly

Both the EV and CR genome were assembled using Falcon, which is specifically designed to assemble diploid genomes using long read libraries (Chin *et al.*, 2016). Length cut-offs for reads to be used by Falcon were chosen to include the longest reads that would give in between 25 and 30X coverage (CR 9kbp; EV 13kbp). Falcon was run with the following sense option settings for both genomes: --output_multi, --min_idt 0.70, --min_cov 5, --local_match_count_threshold 2, and --max_n_read 200. Due to differences in coverage and heterozygosity levels the overlap filtering settings were slightly different between the two with --max_diff 40, --max_cov 60, --min_cov 4, --bestn 10 for EV and --max_diff 50 --max_cov 75 --min_cov 4 --bestn 10 for CR. The BE lineage was recently sequenced (Faddeeva-Vakhrusheva *et al.*, 2017). However, when completing the assembly of its endosymbiont wFol, the scaffolding step with SSPACE was found to have introduced false links between contigs. As genome synteny was important to our question, the three Falcon assemblies were taken and processed in the same way from that starting point. Coverage based trimming was used to filter out any false connections within the contigs from Falcon. Reads were mapped back to the assemblies using BLASR (Chaisson and Tesler, 2012) and regions with a coverage lower than ~10% of the average coverage were filtered out using BEDTools (Quinlan and Hall, 2010). Hereafter, contigs were joined and extended with PBJelly (English *et al.*, 2012). Trimming and PBJelly were repeated a second time and then we concluded with a final trimming step to create a genome without coverage dips below ~10% of the average coverage. Next, the assemblies were polished with Quiver (for both BE and EV lineages) or Arrow (for the CR lineage; both tools from the GenomicConsensus package https://github.com/PacificBiosciences/GenomicConsensus). It was necessary to use two different polish algorithms because of the different output formats from the different PacBio platforms used for sequencing. To identify and filter out Wolbachia and other non-Folsomia contigs, we used Blobtools with a blastn to the nt database, a diamond blast to the Reference_Proteomes_2018_04 database and information on reads mapped back by BLASR for each genome (Laetsch and Blaxter, 2017) as input. To detect possible vector contamination VecScreen was run (National Center for Biotechnology Information, n.d.). Finally, the contigs were reordered to size with a custom python script.

We also extracted, assembled and annotated the mitochondrial genomes (mt genomes) from all three lineages. Reads containing mt DNA were filtered by blasting to the already published mt genome of the BE lineage. These reads were assembled with CANU (Koren *et al.*, 2017). The contig that was build out of most reads was selected and overlapping regions on the sides were trimmed to form one circular contig. Mt genomes were thereafter polished in the same way as their matching *F. candida* genome. These polished genomes were annotated with MITOS2 (Bernt *et al.*, 2013) and checked manually in Artemis (Rutherford *et al.*, 2000).

Annotation

For all three genomes annotation was performed using the MAKER v.2.31.8 pipeline (Holt and Yandell, 2011). First, repeat libraries were made with RepeatModeler (Smit and Hubley, 2008). These were used in combination with the Dfam (Hubley *et al.*, 2016) and Repbase (Bao, Kojima and Kohany, 2015) databases to softmask the genomes by RepeatMasker with RMblast as its search algorithm (Smit, Hubley and Green, 2015). For the BE and EV genomes AUGUSTUS v.3.1.0. (Stanke and Morgenstern, 2005) and GeneMark (Besemer and Borodovsky, 2005) were used as *ab initio* gene predictors trained by BRAKER1 (Hoff *et al.*, 2016) with raw RNA-Seq data of the BE lineage (SRR935329, SRR921597 and SRR1182449). The RNA-Seq data from SRR935329 was first trimmed with Trimmomatic (Bolger, Lohse and Usadel, 2014) and unpaired reverse reads we removed to improve mapping to the genomes. The RNA-seq data of the BE lineage did not share enough similarities with the CR genome. Therefore, the CR genome AUGUSTUS was run using the gene models of the BRAKER run on the BE genome to create the *ab initio* gene predictions. To assess completeness, BUSCO (Simão *et al.*, 2015) was run with Arthropods & Insects v3 gene sets on all genomes and assemblies, as well as on the previous version of the BE genome and annotations for comparison. Functional annotation was assigned with a BlastP search (Altschul *et al.*, 1990) against NCBI's nr database (E-value threshold of 0.1), which is the most complete curated protein database, and InterProScan analysis (Jones *et al.*, 2014) all run with Blast2GO v.5.2.1. To minimize false positive gene identifications, genes under 300bp without annotation were removed.

Phylogenetics

Asexual lineages in *F. candida* can be divided into two phylogenetic clusters (Tully et al. 2006). To assess whether strain BE and EV belong to the same cluster and to clarify the evolutionary history of the sexual CR strain, we extracted 18s and 28s rRNA sequences of our three lineages from their genome assemblies and aligned these with the 18s and 28s rRNA sequences from (Tully et al. 2006). Two

phylogenetic trees were created using the Maximum Likelihood method (Jones, Taylor and Thornton, 1992) in MEGA7 (Kumar, Stecher and Tamura, 2016) following the protocol of Hall 2013 (Hall, 2013). The reliability of the trees was estimated by the Bootstrap method (n=1000) and most suitable substitution models were calculated by MEGA7. This was the Kimura 2-parameter model for 18s and the Jukes-Cantor model for 28s.

Repeat annotation and landscapes

The repeats within the three genomes were annotated separately. Repeat libraries for each lineage were made with RepeatModeler using RMblast as its search algorithm (Smit and Hubley, 2008). These databases in combination with the Dfam (Hubley *et al.*, 2016) and Repbase (Bao, Kojima and Kohany, 2015) databases were used to annotate the genomes by RepeatMasker (Smit, Hubley and Green, 2015). Thereafter, detailed summaries of the RE classes found were created, Kimura distances were calculated and a repeat landscape was visualized by respectively the buildSummary.pl, calcDivergenceFromAlign.pl and createRepeatLandscape.pl scripts from the same package.

Segmental duplications and palindromes

To detect segmental duplications we used the Multiple Collinearity Scan toolkit MCScanX (Wang *et al.*, 2012). The MCScan algorithm detects both synteny and collinearity in genome data and provides several visualisation options. The duplicate gene classifier within MCScanX uses the MCscan algorithm to classify the genomic context of gene duplications. Segmental duplications were classified as duplications of large segments of the genome with at least five matching genes, the rest of the gene duplications with less genes involved were categorized as other duplications. We ran MCScanX with the following parameters: MATCH_SCORE: 50, MATCH_SIZE: 5, GAP_PENALTY: -1, OVERLAP_WINDOW: 5, E_VALUE: 1e-05, MAX GAPS: 25. Collinear blocks of genes on a single chromosome in reverse orientation were marked as palindromes. Locations of palindromes were visualized using Circos (Krzywinski *et al.*, 2009). Similarity between palindromes of the three lineages were determined and visualized with the progressive Mauve algorithm (Darling *et al.*, 2004). A pearson correlation test was performed to test for association between segmental duplications, NUMTs and Res. In addition, a general linearized model with a quasi-poisson distribution was run to test if there was a difference in the number of Res in 100Kb windows with and without segmental duplications and with and without NUMTS of the genome. Only complete windows were used in the analysis, and all statistical analyses were done in R (R Core Team, 2017).

Possible haplotig detection

The Falcon assembler creates a haploid primary assembly of a diploid genome with alternative haplotigs for regions where the two chromosomes differ. To verify that the high numbers of segmental duplications were not caused by haplotigs that remained within the primary assembly, contigs with a high-quality blast alignment to other contigs in the primary assembly for over at least 95% of their length were identified as possible haplotigs (e-value cutoff of 1E-10). Fewer of these possible haplotigs were found in the asexual lineages (BE: 19, EV 22), than in the sexual one (75). However, none of the possible haplotigs in the asexual lineages contained segmental duplications and only 6 out of 75 in the sexual genome did so. Therefore, it seemed unlikely that the segmental duplications could be an artefact created by unfiltered haplotigs (Supplementary table 3). We also accessed the effect filtering these contigs would have on the assemblies by running BUSCO in the same way as to the complete assemblies. BUSCO scores of the three genomes also barely changed (Supplementary figure 2). Moreover, all possible haplotigs were comparatively small and together took up only a very low percentage of the genomes (Supplementary table 3). Thus, because of the low impact and the lack of prove that they were real haplotigs rather than unplaced segmental duplications we decided not to remove these contigs.

Ks analysis of duplications

Gene families of two or more members were compared by measuring the number of synonymous substitutions per synonymous site (Ks value) according to (Vanneste, Van de Peer and Maere, 2013) with some modifications. For each species an all-against-all protein similarity search was done using BLASTP with an e-value cut-off of e^{-10}. From these similarity data, gene families were constructed using Markov Clustering from the MCL pipeline (Enright, Van Dongen and Ouzounis, 2002). The amino acid sequences of gene families were then aligned using MUSCLE (Edgar, 2004). Alignments contained a variable number of gaps, which could influence Ks estimation, and were therefore removed before Ks analysis. Protein families with a gap-stripped alignment length of less than 100 amino acids were not considered for further analysis. The nucleotide sequences of the remaining gene families were then aligned using the protein alignment as a guide in PAL2NAL (Suyama, Torrents and Bork, 2006). For each protein alignment, the Ks value was estimated through maximum likelihood using CODEML from the PAML package (Yang, 2007). Ks values higher than five were filtered out since these are deemed unreliable (Colbourne *et al.*, 2011; Vanneste, Van de Peer and Maere, 2013). The final set of Ks values for each genome is represented as a histogram to detect temporal patterns of gene duplication, with higher Ks indicating older duplications.

Results

Genome assemblies and annotations

For this study we sequenced, assembled and annotated two new F. candida lineages, an asexual population from Evergem, Belgium (EV) and a sexual population from Cremona, Italy (CR). In addition, we partially reassembled and reannotated the previously described asexual Berlin lineage, Germany (BE) to create three comparable genomes with highly confident contiguity. Table 1 gives an overview of the most important assembly statistics for these three genomes. The genome sizes of the two asexual lineages were highly comparable with 220 Mbp for BE and 221 Mbp for EV, and also similar to the first version of the BE genome (BE-v1), which was 221 Mbp (Faddeeva-Vakhrusheva *et al.*, 2017). The genome of the sexual lineage, however, was around 11% smaller with 197 Mbp. The number of contigs also differed, with both asexual lineages having considerable fewer contigs than the sexual lineage (BE: 190, EV: 173, CR: 520). Most likely this is due to the difficulty of assembling the more heterogeneous sequences derived from pooled, non-clonal animals as used for the CR genome. Consequently, the N50 and L50 values were also slightly lower for the CR genome. Nevertheless, with N50s ranging from 2 to 4.7 Mbp these three genomes are among the most contiguous arthropod genomes available. Since we did not allow gaps in the new assemblies and filtered out low coverage areas to prevent chimeric contigs, the N50 is lower than the BE-v1 genome (6.5 Mbp). Thus, this resulted in three high quality assemblies with marginally shorter contigs but higher confidence.

Annotation of the three genomes yielded a considerable number of gene models (Table 1). The asexual lineages had an estimated 43 and 44 thousand gene models, which was higher than the sexual population which had roughly 37 thousand genes predicted. All three assemblies had substantially more genes than found in BE-v1, in which 28,734 gene models were determined. Validation of the genomes based on the presence of Benchmarking Universal Single-Copy Orthologs (BUSCO) showed percentages ranging from 95.7 to 96.2 for all three genomes similar to the 96.2% in BE-v1 (Table 1, Supplementary figure 1). BUSCO scores based on the annotations were even higher (BE: 96.4%, EV: 96.6%, CR: 96.7%; Table 1), suggesting our annotations were better than the non-specific annotation used by BUSCO. In addition, they were also higher than in BE-v1 in which only 79.5% of the genes were retrieved. Hence, the three new assemblies had more complete annotations than the BE-v1 assembly. Thus, a major part of the increase in the number of genes was probably due to the better annotations or an overestimation of the number of genes. The sexual CR lineage had more duplicated BUSCO genes than the two asexual lineages (Supplementary figure 1), which might have to do with the large number of segmental duplications. Also, more than half of the missing BUSCO genes were missing in all three linages, both at genome and

annotation level. They are most likely not present in these genomes at all (Supplementary table 1) and should probably not be seen as essential genes for *F. candida*. Thus, no signs of lack important regions within the assemblies was found.

Table 1: Genomic characteristics of the primary assemblies of F. candida lineages.

	Berlin	**Evergem**	**Cremona**
Assembly			
Number of bases (Mbp)	220	221	197
Number of contigs	190	173	520
Min sequence length (bp)	995	1215	1130
Max sequence length (Mbp)	18.4	20.5	18.6
N50 (Mbp)	3.5	4.7	2.0
L50	15	12	19
GC%	37.5	37.5	36.8
Structural annotation			
Number of genes	43065	43938	37632
Number of transcripts	45814	46593	37632
Alternative splice variants	2749	2655	0
Functional annotation			
Interpro	34137	34557	29704
nr database	33650	33464	26123
GO mapping	26444	26336	20350
B2G Annotated	21499	20993	13022
Validation			
BUSCO % complete in genome	1024 (96.1%)	1020 (95.7%)	1025 (96.2%)
BUSCO % complete in annotation	1028 (96.4%)	1030 (96.6%)	1031 (96.7%)
Repeats			
DNA	4.08%	3.87%	2.21%
LINE	0.72%	0.80%	1.30%
LTR	0.84%	0.78%	1.79%
SINE	0.01%	0.03%	0.03%
Helitron	0.76%	1.08%	0.62%
Unknown	13.76%	13.95%	8.67%
Total interspersed	20.16%	20.52%	14.58%
Low complexity	0.12%	0.12%	0.14%
Satellite	NA	0.03%	0.00%
Simple repeats	1.13%	1.15%	1.30%
Total percentage of genome	21.42%	21.83%	16.02%
Total number of repeats	237679	234462	128991
Total number of bases (Mbp)	47.1	48.2	31.6
Segmental duplications			
Between contigs	91	100	232
Tandem	8	5	4
Palindromes	16	16	17
Total number of SDs	115	121	253
Total number of genes included	1358	1519	5883

Mitochondrial genomes and NUMTs

Complete mitochondrial genomes (mt genomes) were assembled and annotated for all three *F. candida* lineages (Figure 1). All three lineages contained the full set of 37 mitochondrial genes and were around 15,000 base pairs long (BE: 15,178, EV: 15,186, CR: 14,987; Table 2). In addition, we found signs of multiple nuclear

mitochondrial DNA segments (NUMTs) in each of the three genomes, with the sexual CR lineage containing more NUMTs than the two asexual lineages (Table 2). However, long NUMTs (>1000 bp) were more numerous in the two asexual lineages, leading to higher numbers in total bp. The largest NUMTs in the BE and EV genomes were 13,894 base pairs and 14,875 respectively, suggesting close to full-length transfers of the mt genome into the nuclear DNA (Table 2).

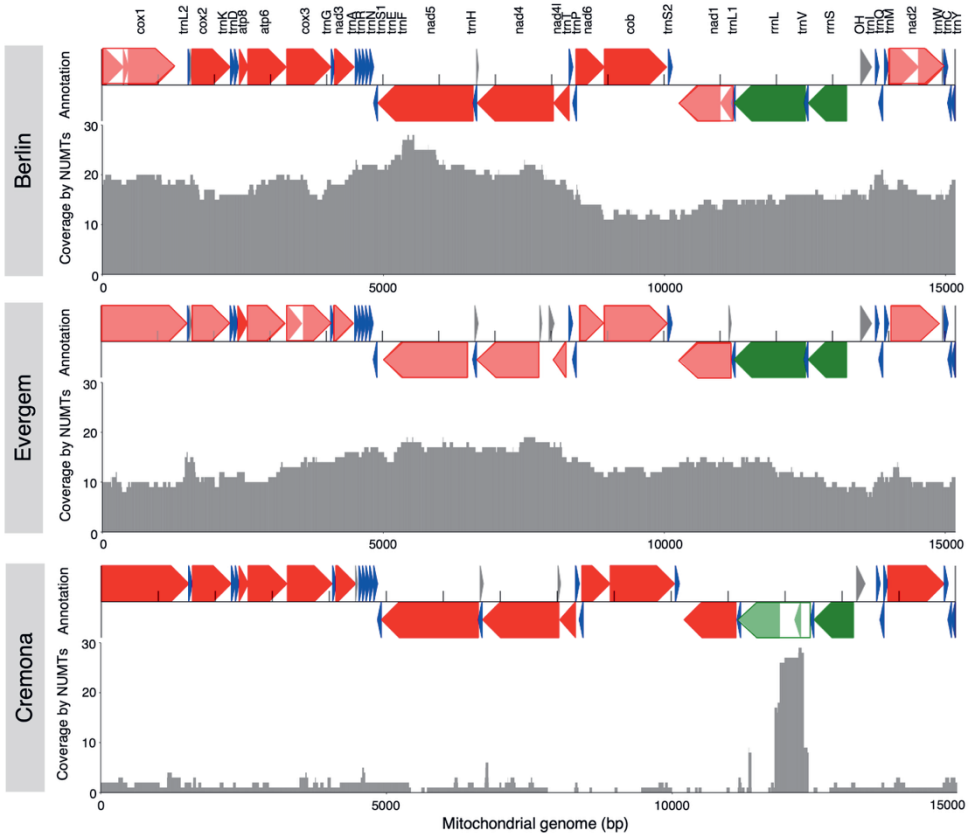


Figure 1: Schematic representations of mitochondrial genome annotations and histograms of NUMT coverage over these genomes per *F. candida* lineage. The 22 tRNA genes are blue, the 13 protein-coding genes red, the 2 rRNA genes green and the origin of replication of the heavy strand (OH) in grey. Genes that were indicated as split or containing an internal stop codon are display in a lighter colour with a full colour outline. Coverage by NUMTs was based on overlapping blast alignments. Annotations were done by MITOS2.

Multiple copies of the entire mt genome were represented by NUMTs of both asexual lineages (BE: 11X and EV: 7 X; Table 2 and Figure1). By contrast, NUMTs in the CR genome did not even cover the whole mt genome once. Instead, a peak in coverage was found at the 16S ribosomal RNA gene (rrnL, Figure 1). Several genes in the mt genome included internal shifts or stop codons, characterizing them as

pseudogenes that possibly lack proper functionality (Figure 1). Both asexual linages contain multiple pseudonised genes, while the sexual lineage only contains one split gene (rrnL) at the peak in coverage previously mentioned. Since all mitochondrial genes are necessary for survival of animals, pseudogenization of mitochondrial genes requires intact versions of the genes in the NUMTs to have taken over their functions. Alternatively, the inferred mitochondrial pseudogenes could be the result of assembly errors that resulted from chimeric contigs from NUMTs and mitochondrial pieces of the genomes.

Table 2: Characteristics of mitochondrial genomes and NUMTs of F. candida lineages.

	Berlin	**Evergem**	**Cremona**
Mitochondrial genome			
Length (bp)	15,178	15,186	14,987
Genes annotated	37	37	37
NUMTs			
Number	71	64	102
> 1000 bp	43	34	4
Longest	13,894	14,875	5,712
Shortest	42	40	41
Total bp	265,997	198,488	31,253
Density (bp/kb)	1.208	0.898	0.159
Min coverage (X)	11	7	0

Phylogenetics

Phylogenies were made based on two ribosomal markers to determine the evolutionary relationship and relatedness of the BE, EV and CR linages to each other and additional *F. candida* lineages (Supplementary figure 2). Consistent with previous work, two clusters were present within the asexual lineages of *F. candida* (Tully *et al.*, 2006). Our asexual strains EV and BE were both with the B cluster. Both trees indicated that the BE and EV lineages are most closely related and that the CR population is a sister group to them.

Repeat annotation and landscapes

RepeatMasker was used to quantify and annotate repeats and to assess the temporal pattern in which repeats duplicate within the genomes (Table 1, Supplementary table 2). Clear differences exist between the sexual CR and the two asexual lineages, with the CR genome containing only 16.0 percent of repeats, while BE and EV respectively contain 21.4 and 21.8 percent. Most of the Res that were identified in the *F. candida* genomes were unclassified repeats, which is typical for arthropod genomes and is probably caused by the lack of functional studies into their mobilomes (Petersen *et al.*, 2019). Of the classified repeats, most were interspersed repeats. Of the major classes within these interspersed repeats,

the sexual genome contains more LINE and LTR repeats and fewer DNA and unclassified repeats than the asexual genome. By far the largest number of interspersed repeats in all three genomes could not be classified and were therefore characterized as unknown. The repeat landscapes also showed clear differences between the sexual and the two asexual lineages (Figure 2). The sexual population had a very high number of recent repeats followed by a steady decline in number of older repeats, while the asexual lineages had a less pronounced peak of recent repeats and a second peak of older repeats with a Kimura substitution level of around 12. This second peak suggests that there has either been a temporary increase in repeat duplication activity in ancient time or a more recent decrease in duplication rate.

Segmental duplications and palindromes

To further investigate the dynamics with the *F. candida* genomes, we searched for segmental duplications using MCScanX. Large numbers of segmental duplications were found in all three genomes, with the CR genome containing most duplications (BE: 115, EV: 121, CR: 253; Table 1). Most of the segmental duplications were on separate contigs, but several tandem segmental duplications were also found. In addition, significant numbers of palindromes were found in all three *F.candida* genomes (BE: 16, EV: 16, CR: 17; Table 1). The palindromes were distributed over 11 to 13 contigs and in some cases multiple palindromes were located on one contig (Figure 3A). Comparing the set of palindromes between the three lineages showed that roughly one third of the palindromes of the asexual lineages shared regions of high similarity (Figure 3B). Remarkably, the palindromes within the sexual CR genome shared no regions of high similarity with those in the two asexual genomes.

Focusing on the location of NUMTs and segmental duplications it becomes apparent that they are overrepresented in the windows with relatively many REs, as was visualised by plotting the number of REs in 100Kb windows with and without these genomic features (Figure 4). For all three lineages, most distributions of the REs within the 100Kb window follows a bimodal distribution with a long tail of outliers. The two peaks represent 100Kb windows with relatively few and with relatively many REs. For the segetal duplications, the two peaks are further apart in the asexual lineages than in the sexual lineage, suggesting a stronger separation between the two classes of windows. For areas with or without NUMTs the separation is also visible, but the distributions do not form clear unimodal peaks. Especially, for the distribution of REs in 100 Kb areas with NUMTs in the sexual lineage this is the case.

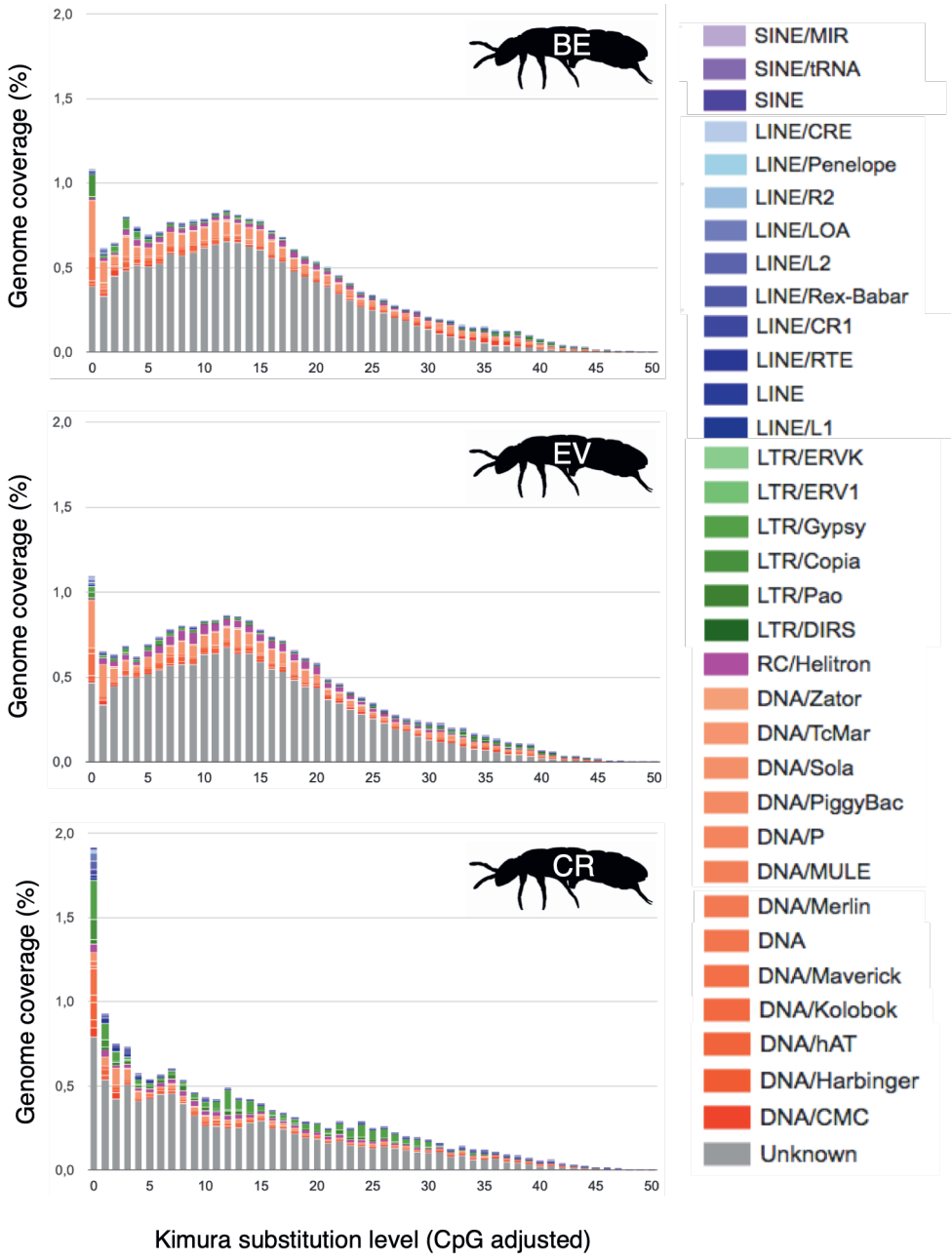


Figure 2: Repeat landscapes plots of the three *F. candida* lineages, which display the approximate age (Kimura substitution level) and coverage of the genome by repeats of different families. Names of the lineages are abbreviated the two asexual lineages Berlin and Evergem as BE and EV and the sexual Cremona population as CR.

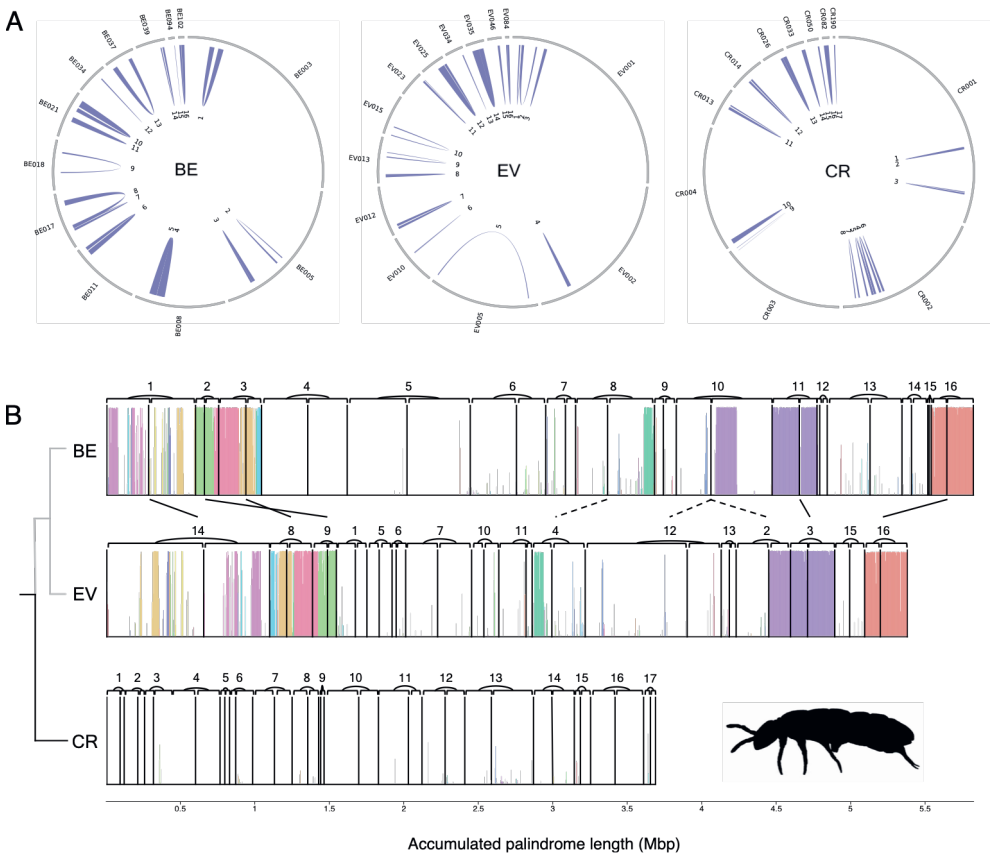


Figure 3: Locations and similarities of palindromes in the F. candida genomes. A) Circular maps of the locations of all palindromes in the three F. candida lineages. Only contigs that contain palindromes are displayed, they are numbered in accordance with the assembly. B) Overview of the similarities between the palindromic regions within the three F. candida genomes. On the left is a simplified phylogeny of the F. candida lineages based on the phylogenies in supplementary figure 2. Matching coloured areas indicated that these regions are similar and the height of the bars indicates the level of similarity. Palindromes are numbered according to the MCScanX output in both graphs. Names of the lineages are abbreviated the two asexual lineages Berlin and Evergem as BE and EV and the sexual Cremona population as CR.

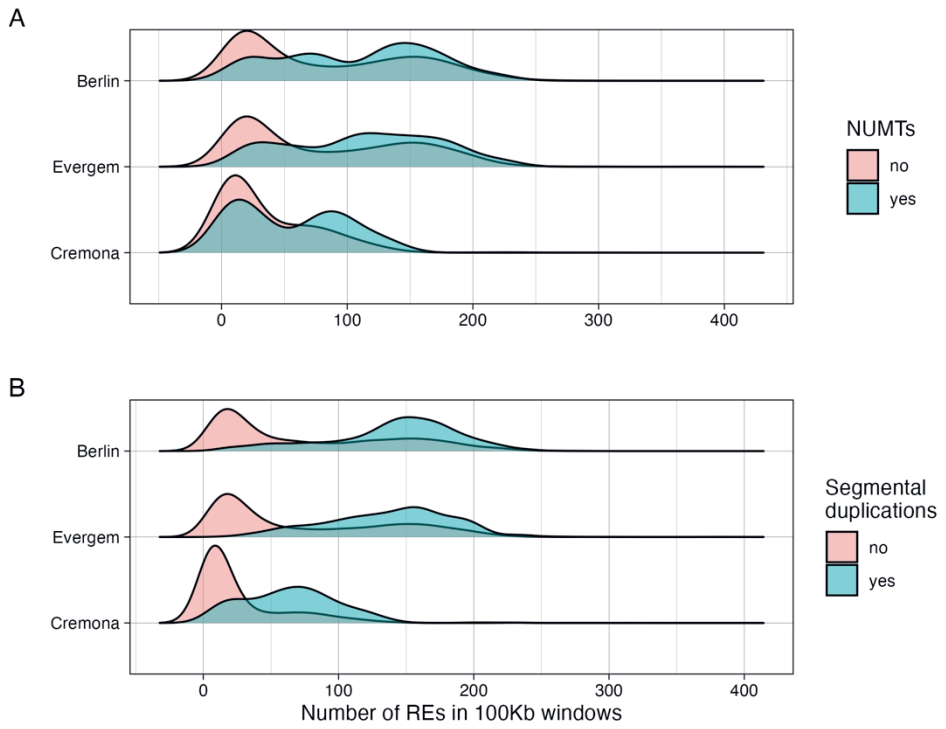


Figure 4: Ridge plots of the number of repetitive elements (REs) within 100 Kb windows of the genomes with or without NUMTs (A) and segmental duplications (B). For the number of REs in combination with NUMTs we found significant differences between presence and absence ($F(1) = 28.48$, $p < 0.001$) and between the lineages ($F(2) = 461.07$, $p < 0.001$) with the sexual lineage significantly deviating from both asexual lineages. For the number of REs in combination with segmental duplications, we found significant differences between presence and absence ($F(1) = 676,69$, $p < 0.001$), between the lineages ($F(2) = 480.41$, $p < 0.001$) and a significant interaction between lineages and presence ($F(2) = 9.73$, $p < 0.001$).

Ks analysis of duplications

To determine the age of segmental duplication, we calculated the Ks values per gene pair within segments and compared them to all other gene duplications within the *F. candida* genomes. This resulted in a similar pattern for the three genomes, in which there were many recent segmental duplications followed by a long tail of older ones (Figure 5). For BE, the curve was less smooth than for the other two, suggesting that the duplications occurred in small bursts rather than continuously. Furthermore, the sexual CR lineage had a much higher peak with low Ks values, which implies it had a faster rate of segmental duplications. Finally, based on the distributions, the retention of intact segmental duplications seems to be less than that of other duplications, probably because they break into smaller pieces.

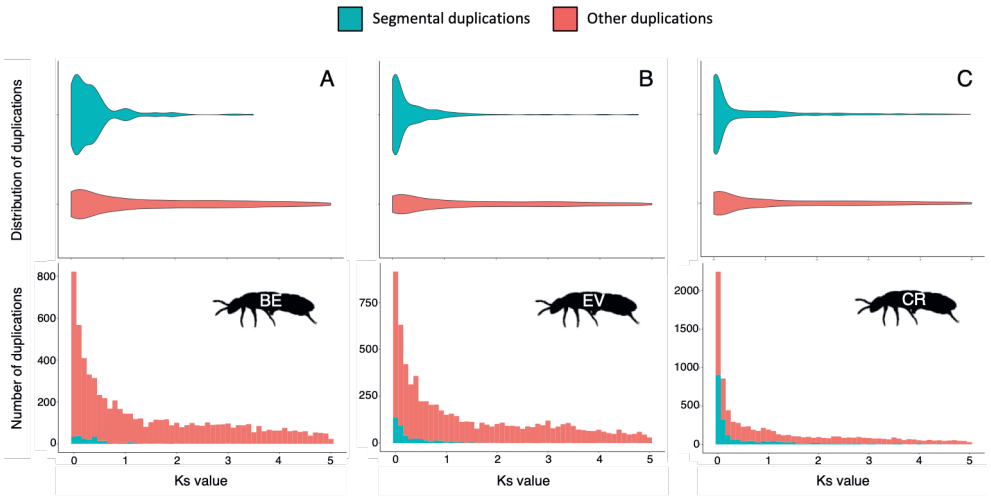


Figure 5: Ks plots of segmental and other smaller duplications within the *F. candida* genomes. On the top panels the distribution of the duplications and on the bottom the number of duplications per lineage. Names of the lineages are abbreviated, the two asexual Berlin (A) and Evergem (B) as BE and EV and the sexual Cremona (C) population as CR.

Discussion

Genomes of the three lineages of *F. candida* springtails with different modes of reproduction were characterized and compared to assess the differences in genome dynamics between sexual and asexual lineages. The genome of the sexual lineage was around 11% smaller and had around 14 % fewer genes than the asexual lineages. Differences were also found in the number of duplications for both repeats and SDs, with the sexual lineage having fewer repeats but more SDs than both asexual genomes. Moreover, a change in the rate of duplications and repeats was apparent in the asexual ages, whereas the sexual lineage contained large numbers of new duplications that degraded rather fast for both repeats and SDs. We also found big differences in the amount of mitochondrial DNA in the form of NUMTs, as the asexual lineages had long NUMTs that covered the complete MT genome multiple times, while the sexual genome had many short NUMTs that did not fully cover the MT genome. Both SDs and NUMTs were found in areas with high RE densities suggesting that these regions are more dynamic than the rest of the genome. Finally, palindromes were found in all three of the genomes in equal numbers, though they were dissimilar in sequence between the sexual and the asexual lineages.

Differences in duplication events and rates

One of the key factors influencing genome dynamics is the number and retention of duplications (Levasseur and Pontarotti, 2011). Theory predicts that lack of recombination will lead to the accumulation of Res and other repetitive DNA (Lynch and Blanchard, 1998; Bachtrog, 2013; Leung *et al.*, 2015), because lack of recombination makes it more difficult to eliminate Res from the genome (Nuzhdin and Petrov, 2003). Hence asexual linages are expected to accumulate many Res resulting in expansion and qualitative degradation of their genomes. Consistent with theory, our results clearly showed that asexual lineages contained a higher number of Res than the sexual.

High numbers of segmental duplications were found in the previous version of the BE genome (Faddeeva-Vakhrusheva *et al.*, 2017). Here, we confirm this finding and also report high numbers of segmental duplication in the other two *F. candida* genomes. Moreover, like what we found for the other repeats, a recent drop in duplication activity for the asexual lineage was evident for SDs. The sexual lineage of *F. candida*, however, contained even more segmental duplications than both asexual lineages.

Finding segmental duplications in all three genomes is remarkable, as there are few arthropod genomes that contain these elements (Roelofs *et al.*, 2020) and could potentially cause genomic instability as they can lead to chromosomal rearrangements (Emanuel and Shaikh, 2001). Apparently, the extraordinary conditions that led to this large number of segmental duplications were already in place in the common ancestor of the sexual and asexual lineages. A possible explanation for such a high number of SDs would be a whole genome duplication (WGD) which has been suggested for *F. candida* (Li et al. 2018). However, the presence of tandem SDs and palindromes that are on the same chromosome are not consistent with WGD (Roelofs *et al.*, 2020). Moreover, WGD would appear as a hump in the Ks plot, revealing a certain point in time at which the duplication happened. Since we find tandem SDs, palindromes and a Ks distribution that suggest a continuous generation of new SDs in all three genomes, it is very unlikely that this was caused by a WGD event. Thus, there must be another mechanism at play that causes these high rates of duplications of both repeats and SDs that is yet to be defined.

NUMTs (Integration of mitochondrial DNA)

With NUMT densities of 1.208 bp/kb and 0.898 bp/kb the BE and EV lineages of F. candida are the highest and third highest in any animal studied so far. For example, densities in the human genome are six times lower (Tao *et al.*, 2023). Even the honeybee genome, which contains more than 800 relatively small NUMTs and has been recognized for its exceptionally high density of NUMTs, density only just

surpasses 1.0 bp/kb (Pamilo, Viljakainen and Vihavainen, 2007). These high densities in the honeybee are most likely caused by secondary transposition, as two-thirds of the NUMTs are flanked by 'mariner' type transposons (Pamilo, Viljakainen and Vihavainen, 2007). This might also be the case for the NUMTs in the asexual *F. candida* lineages. However, the NUMTs in these lineages are much longer than the NUMTs in the honeybee, suggesting fewer integration events in combination with good preservation of these NUMTs. Moreover, the low density of NUMTs in the sexual lineage 0.159 bp/kb, suggests that there was an increase in NUMTs after the transition to asexual reproduction. Alternatively, the integration of NUMTs in the genome could have aided in becoming a successful asexual lineage. As it was hypothesized that the mismatch in mutation speed of the mitochondrial genome with the nuclear genome favoured the evolution of sexual reproduction (Havird, Hall and Dowling, 2015), the integration of mitochondrial genes within the genome would offer more favourable conditions for asexual reproduction. In the case that the NUMTs take over the mitochondrial functions, the mitochondrial genome would start to degrade under its high mutation rate. Also, this degradation was seen in the mitochondrial genomes of the asexual lines. Together, it suggests the NUMTs could be a key part of the success of these asexual lineages, although whether this was a prerequisite or a consequence of this transition has yet to be determined.

Palindromes

Palindromes were not only present in the asexual lineages, but also in the sexual lineage. It has been hypothesized that palindromes are generally found in regions that lack recombination like sex-chromosomes or genomes of asexual species (Rozen *et al.*, 2003; Bachtrog, 2013; Flot *et al.*, 2013; Nowell *et al.*, 2018), although they have not been found in all asexual genomes (Jaron *et al.*, 2020). We detected palindromes in both sexual and asexual lineages of *F. candida* and can therefore conclude that they are certainly not consequence of asexuality in the case of *F. candida*. Furthermore, the palindromes we found were spread over multiple large contigs, suggesting that these are not all located on the same chromosome, making it even less likely that the palindromes are located on a sex chromosome. Instead, the lack of similarity in palindromes between the sexual and asexual lineages and the limited similarity between the asexual lineages, suggest a continuous formation and disappearance of new palindromes over a time span shorter than the split between sexual and asexual lineages, indicating that these genomes are very dynamic on the larger scale.

Genome dynamics

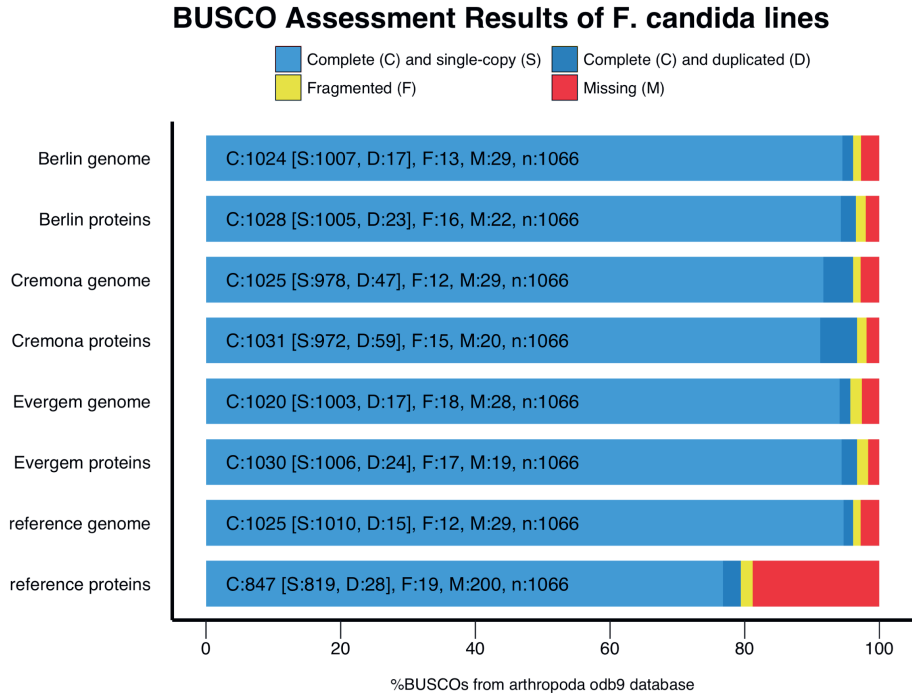
Overall, both sexual and asexual *F. candida* genomes were characterized by high density of both mitochondrial DNA in the form of NUMTs and duplicated DNA in the form of REs and segmental duplications. However, the genome dynamics causing these high densities seem to be very different between sexual and asexual lineages. In the asexual lineages this may be due to better conservation of these elements, while the sexual lineage could be more susceptible to infiltration, as suggested by the higher number of recent elements. The larger genome size and number of genes in the asexual lineages is most likely caused by this shift in dynamics from purging to conservation that resulted in an expanded genome retaining much of the new genetic material coming in. In addition, NUMTs and segmental duplications are more numerous in segments of the genome with high RE concentrations, indicating that these areas are less stable. The separation between these less stable areas with high RE concentrations and the rest of the genome is more pronounced in the asexual lineages, suggesting that within these genomes there are regions that are under stricter control than the rest, a pattern known as the two-speed genome model also seen in plant pathogenic fungi genomes (Torres *et al.*, 2020). Separating dynamic regions from conserved regions may help asexual lineages to survive as it limits degradation of important regions. Intriguingly, we observed a similar pattern of high conservation of DNA in the *Wolbachia* endosymbiont of the asexual lineages of *F. candida*, *w*Fol (Kampfraath *et al.*, 2019). The *w*Fol genome was larger and contained more intact REs and phages than any other *Wolbachia.* Moreover, it displayed regions with and without strong conservation. The parallel patterns observed in these two unrelated genomes may be due to similarities in evolutionary pressures. Becoming asexual and becoming restricted within one host both led to smaller effective population sizes for parthenogenetic lineages of *F. candida* and *w*Fol. A smaller population size has been theorized to result in a larger genome size (Lynch and Conery, 2003) and found when less effective selection due to smaller population size resulted in larger genomes of asellid isopods when comparing 11 independent habitat shifts (Lefébure *et al.*, 2017). The reduction in effective selection of *F. candida* and *w*Fol might have also led to a situation in which it became more difficult to repair or replace damaged or lost genes. Leading to a situation in which only genomes that were able to preserve a working set of genes would survive.

Conclusions

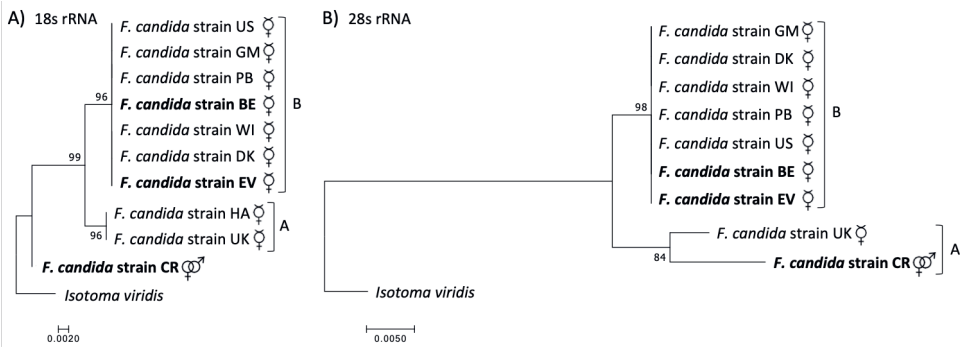
Our comparative genomic analysis revealed that both sexual and asexual lineages possess palindromic regions, portraying this is not a sign of asexuality in *F. candida*. However, profound differences in genome architecture between the sexual and asexual lineages of *F. candida* were revealed. While the sexual lineage possesses a

smaller and more streamlined genome, the asexual lineages exhibit signs of genetic expansion, characterized by significantly elevated levels of repetitive elements (REs), segmental duplications (SDs), and mitochondrial DNA remnants (NUMTs). This pattern suggests that transitioning to an asexual mode of reproduction is associated with a unique genomic strategy: one that promotes both accumulation and retention of diverse genetic material. Specifically, the ability of asexual lineages to tolerate and retain large amounts of repetitive DNA (REs, SDs) and integrate mitochondrial material (NUMTs) may represent an adaptive evolutionary strategy to overcome need of sexual recombination.

Supplementary material



Supplementary figure 1: BUSCO results of the three F. candida lineages and the reference (previous version of the Berlin lineage) to the arthropoda odb9 database.



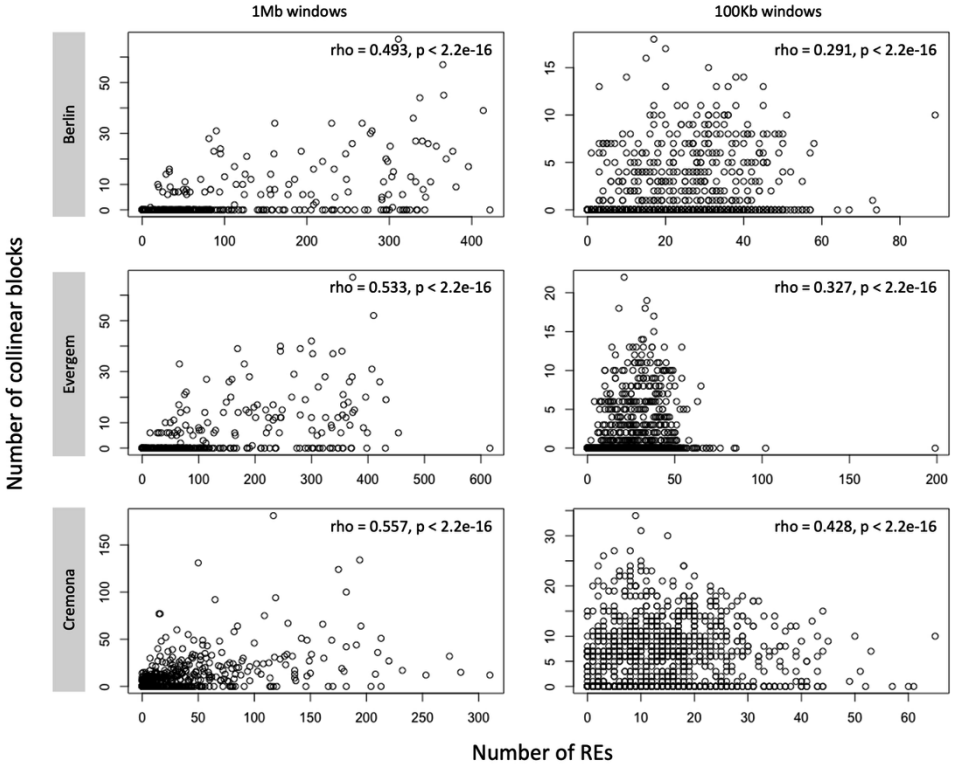
Supplementary figure 2: Maximum Likelihood phylogenies of *F. candida* lineages with *Isotoma viridis* as an out-group. A) Of 18s rRNA genes generated using the Kimura 2-parameter model based on 571 nucleotide positions left after allowing fewer than 5% alignment gaps, missing data, and ambiguous bases at any position. B) Of 28s rRNA genes of *F. candida* lineages with *I. viridis* as an outgroup, generated using the Jukes-Cantor model based on 506 nucleotide positions left after allowing fewer than 5% alignment gaps, missing data, and ambiguous bases at any position. Both trees are drawn to scale in MEGA7 with branch lengths measured in the number of substitutions per site generated and Bootstrap percentages next to the branches are based on 1000 replicates.

Supplementary table 1: Missing BUSCOs per lineages in both genome assembly and annotation. Genes missing in all columns are coloured in grey.

BE genome	EV genome	CR genome	BE proteins	EV proteins	CR proteins
EOG090X00VK	EOG090X00VK	EOG090X00VK	EOG090X00VK	EOG090X00VK	EOG090X00VK
EOG090X01H4	EOG090X01H4	EOG090X01H4	EOG090X01H4	EOG090X01H4	EOG090X01H4
EOG090X02LG	EOG090X02LG			EOG090X02LG	EOG090X02LG
		EOG090X02LT			EOG090X02LT
			EOG090X0398	EOG090X0398	
EOG090X0464	EOG090X0464	EOG090X0464	EOG090X0464	EOG090X0464	EOG090X0464
		EOG090X04G9			
EOG090X04IO	EOG090X04IO	EOG090X04IO	EOG090X04IO	EOG090X04IO	EOG090X04IO
			EOG090X06SH	EOG090X06SH	
EOG090X06HD	EOG090X06HD				
EOG090X06RJ	EOG090X06RJ				
		EOG090X04L0			
EOG090X06YP		EOG090X06YP			
EOG090X076T	EOG090X076T	EOG090X076T			
EOG090X07HG	EOG090X07HG	EOG090X07HG	EOG090X07HG	EOG090X07HG	EOG090X07HG
EOG090X07V0	EOG090X07V0	EOG090X07V0			
EOG090X094T	EOG090X094T	EOG090X094T	EOG090X094T	EOG090X094T	EOG090X094T
			EOG090X09WX		
EOG090X0A9P	EOG090X0A9P	EOG090X0A9P			
EOG090X0AGU	EOG090X0AGU	EOG090X0AGU	EOG090X0AGU	EOG090X0AGU	EOG090X0AGU
			EOG090X0BXN	EOG090X0BXN	EOG090X0BXN
EOG090X0ASF	EOG090X0ASF	EOG090X0ASF			
EOG090X0BAY	EOG090X0BAY	EOG090X0BAY			
EOG090X0BFL	EOG090X0BFL				
	EOG090X0BW7				
		EOG090X0BNZ			
EOG090X0CJ3	EOG090X0CJ3	EOG090X0CJ3	EOG090X0CJ3	EOG090X0CJ3	EOG090X0CJ3
			EOG090X0D2W	EOG090X0D2W	
			EOG090X0EJQ		
			EOG090X0F2L		
		EOG090X0CTK			EOG090X0CTK
					EOG090X0EO1
EOG090X0EPT	EOG090X0EPT	EOG090X0EPT			
EOG090X0FOW					
EOG090X0FVK	EOG090X0FVK	EOG090X0FVK	EOG090X0FVK	EOG090X0FVK	EOG090X0FVK
EOG090X0GME	EOG090X0GME	EOG090X0GME			
		EOG090X0HB3			
					EOG090X0HJH
EOG090X0HLW	EOG090X0HLW	EOG090X0HLW	EOG090X0HLW	EOG090X0HLW	EOG090X0HLW
			EOG090X0I3Y		
		EOG090X0J87			
EOG090X0JQ4	EOG090X0JQ4	EOG090X0JQ4			
EOG090X0KMN	EOG090X0KMN	EOG090X0KMN	EOG090X0KMN	EOG090X0KMN	EOG090X0KMN
EOG090X0LZH	EOG090X0LZH	EOG090X0LZH	EOG090X0LZH	EOG090X0LZH	EOG090X0LZH
			EOG090X0MNX	EOG090X0MNX	
EOG090X0N0O	EOG090X0N0O	EOG090X0N0O			
					EOG090X0OUK
			EOG090X0PCK	EOG090X0PCK	EOG090X0PCK
EOG090X0QLW	EOG090X0QLW				
EOG090X0RU6	EOG090X0RU6				

Supplementary table 2: Repeats per *F. candida* lineage annotated to the family level.

		Berlin			Evergem			Sexual		
Class	Familly	Count	bpMasked	% masked	Count	Bp Masked	% masked	Count	Bp Masked	% masked
DNA	CMC-Chapaev							526	155562	0.08%
	CMC-EnSpm	3011	1147545	0.52%	3737	951102	0.43%	2422	499471	0.25%
	Crypton-C	27	4690	0.00%						
	Ginger				37	16434	0.01%			
	Kolobok-Hydra						45	109629	0.06%	
	MULE-MuDR	31	1660	0.00%	82	7785	0.00%	102	14650	0.01%
	MULE-NOF	256	119394	0.05%						
	Maverick	325	512107	0.23%	342	530554	0.24%	165	414595	0.21%
	Merlin							218	75114	0.04%
	MuLE-MuDR	235	55813	0.03%	591	163148	0.07%			
	MuLE-NOF							36	58052	0.03%
	P	50	50916	0.02%						
	PIF-Harbinger	407	130898	0.06%				361	294910	0.15%
	PIF-ISL2EU							69	70069	0.04%
	PIF-Spy	194	77602	0.04%	68	54689	0.02%	282	142630	0.07%
	PiggyBac	63	42501	0.02%	89	41858	0.02%	91	39682	0.02%
	Sola	75	88107	0.04%	255	25142	0.01%			
	Sola-2				133	53595	0.02%			
	TcMar							62	13270	0.01%
	TcMar-Fot1	28	14170	0.01%						
	TcMar-Mariner	470	318851	0.14%	770	417094	0.19%	222	179950	0.09%
	TcMar-Pogo	61	12105	0.01%						
	TcMar-Tc1	5986	4122386	1.87%	5824	4054475	1.84%	1834	1394507	0.71%
	TcMar-Tc2				121	12883	0.01%	35	7066	0.00%
	TcMar-Tc4	1190	387524	0.18%	345	150638	0.07%			
	TcMar-m44	31	113925	0.05%	57	91323	0.04%			
	Zator	2418	299575	0.14%	2274	289360	0.13%	412	80207	0.04%
	hAT	221	81392	0.04%	727	109895	0.05%	129	112980	0.06%
	hAT-Ac	2700	699689	0.32%	4878	1003867	0.45%	866	385718	0.20%
	hAT-Blackjack	136	29513	0.01%	228	30759	0.01%			
	hAT-Charlie	239	47102	0.02%				95	36159	0.02%
	hAT-Tag1	106	27321	0.01%						
	hAT-Tip100	355	76548	0.03%	370	65245	0.03%			
	hAT-hAT19	316	49471	0.02%	259	74910	0.03%			
	hAT-hATx	86	35737	0.02%	187	61861	0.03%	227	62738	0.03%
	hAT-hobo	347	49084	0.02%				399	79566	0.04%
	Unspecified	1438	399719	0.18%	1189	367568	0.17%	139	103145	0.05%
total DNA		20802	8995345	4.08%	22563	8574185	3.87%	8737	4329670	2.21%
LINE	CR1							93	85537	0.04%
	CRE				18	3864	0.00%			
	CRE-II	488	369890	0.17%	687	578457	0.26%	817	890715	0.45%
	I-Jockey	38	5136	0.00%				132	38817	0.02%
	L1	143	148167	0.07%				79	52127	0.03%
	L1-Tx1				249	155389	0.07%	166	213060	0.11%
	L2	740	324053	0.15%	1278	421108	0.19%	754	626142	0.32%
	LOA	369	167055	0.08%	76	107422	0.05%	292	155448	0.08%
	Penelope	848	332594	0.15%	335	200859	0.09%	493	290417	0.15%
	R1-LOA				67	33036	0.01%			
	R2							60	13097	0.01%
	R2-NeSL	16	20269	0.01%	22	35940	0.02%	101	36297	0.02%
	RTE-X	259	105662	0.05%	339	171501	0.08%	74	71244	0.04%
	Rex-Babar	119	82514	0.04%	139	66966	0.03%	29	39563	0.02%
	Unspecified							36	14081	0.01%
total LINE		3020	1555340	0.72%	3210	1774542	0.80%	3126	2526545	1.30%
LTR	Copia	773	335761	0.15%	216	133360	0.06%	410	430245	0.22%
	DIRS	279	83811	0.04%	467	274673	0.12%			
	ERV1	76	23392	0.01%						
	ERVK	26	14426	0.01%						
	Gypsy	2214	877892	0.40%	1515	701892	0.32%	1344	2058019	1.05%
	Pao	331	505980	0.23%	710	622861	0.28%	824	1017032	0.52%
total LTR		3699	1841262	0.84%	2908	1732786	0.78%	2578	3505296	1.79%
RC	Helitron	6292	1662090	0.76%	9285	2389549	1.08%	3461	1218709	0.62%
SINE	ID				72	11703	0.01%			
	MIR	161	28592	0.01%	107	13758	0.01%			
	tRNA				70	14836	0.01%			
	tRNA-Meta	61	10616	0.00%						
	Unspecified							160	62149	0.03%
total SINE		222	39208	0.01%	249	40297	0.03%	160	62149	0.03%
Unknown		142256	30269245	13.76%	135008	30812594	13.95%	52068	17066165	8.67%
total interspersed		176291	44362490	20.16%	173223	45323953	20.52%	70130	28708534	14.58%
Low complexity repeats		5816	273569	0.12%	5824	273614	0.12%	6191	282406	0.14%
Satellite	5S							35	6131	0.00%
	Unspecified				139	61145	0.03%			
Simple repeats		55572	2489502	1.13%	55276	2550111	1.15%	52635	2559578	1.30%
Total		237679	47125561	21.42%	234462	48208823	21.83%	128991	31556649	16.02%



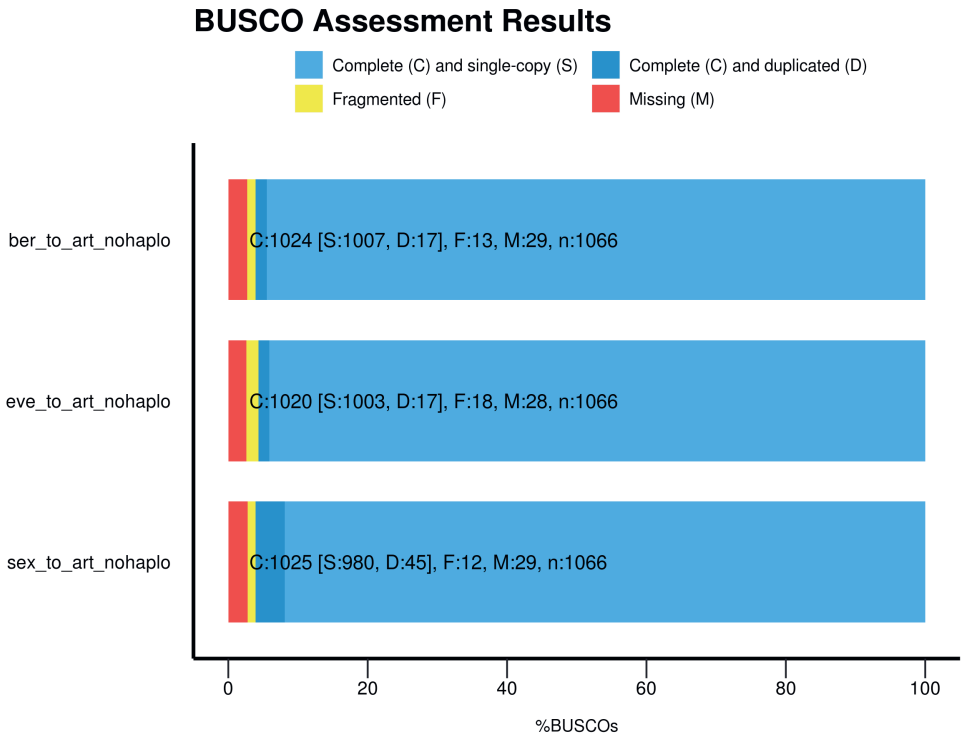
Supplementary figure 3: Correlations between the number of genes from segmental duplications and the number of repetitive elements (REs) within 1Mb and 100Kb windows of the genomes of the three *F. candida* lineages.

Supplementary table 3: Data and mapping stats of the three *F. candida* lineages.

	Berlin	**Evergem**	**Cremona**
Sequence data			
Platform	PacBio RSII	PacBio RSII	PacBio Sequel
Nr. reads	2,003,644	2,057,289	5,552,683
Average length	8,290	9,388	7,691
Shortest	35	35	50
Longest	83,242	78,664	122,900
Total (Mbp)	16,611	19,314	42,708
Minimap2 mapping			
X coverage	65±41	76±78	196±170
mapping %	96.53	96.24	94.52

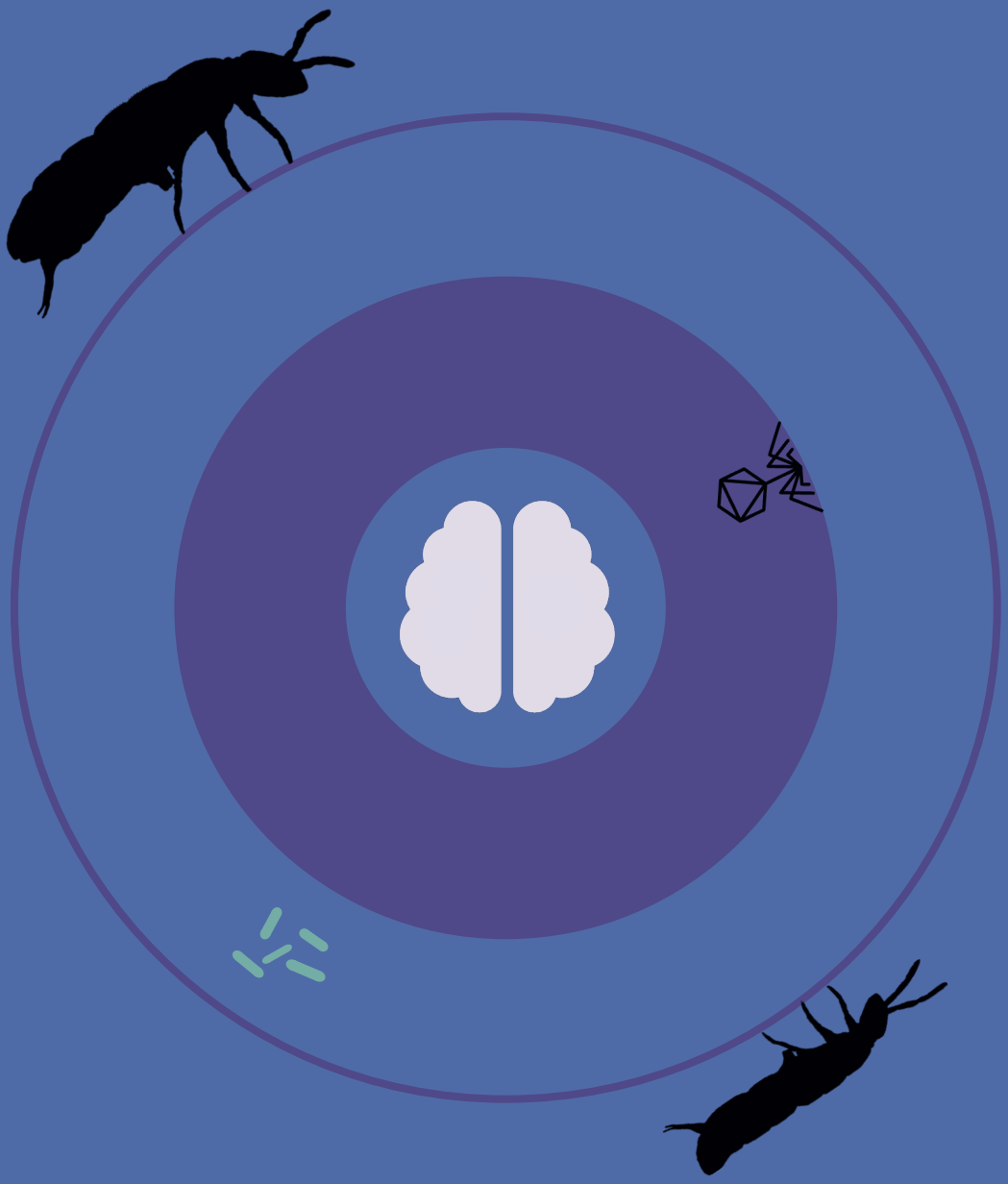
Supplementary table 4: Possible haplotigs characteristics for the three *F. candida* assemblies.

	Berlin	Evergem	Cremona
Haplotigs			
number	19	22	75
with segmental dup.	0	0	6
longest contig	17016	26549	33353
total bp	144492	227219	927998
% of genome	0.06	0.10	0.47



Supplementary figure 4: BUSCO results after possible haplotig filtering.

CHAPTER 6



DKJR

General discussion

This thesis focusses on the evolutionary consequences of an endosymbiont-associated switch from sexual to asexual reproduction and asks whether this results in compensated trait loss within the *Folsomia candida* and *w*FoI study system. For that purpose, male sexual traits, transinfections, and genomic characteristics of both organisms were studied. First, it was demonstrated that several sexual traits of asexual *Folsomia candida* lineages are degenerated. Second, the genomes of the asexual linages of *F. candida* are larger compared to that of the sexual linage, which is probably caused by the accumulation of duplicated DNA over time. Third, we discovered that *w*FoI is the largest *Wolbachia* genome, because of the accumulation of multiple large prophage regions. Through experimental work we also revealed that these prophage regions do not become lytic under heat stress and could confirm that heat stress and antibiotic have direct negative effects on *Wolbachia* titer within *F. candida*, leading to unhatched eggs. These new findings have led to new hypotheses about the dynamics and effects of *Wolbachia* infection, of which the most prominent ones will be addressed in this final part of the thesis, the general discussion.

(Un)stable *Wolbachia* infections

One of the goals of this thesis was to understand the alterations induced by endosymbiont infection. The transinfection experiment in Chapter 2 yielded useful discoveries but unfortunately failed in creating a stable infection in the sexual *F. candida* line by injecting juvenile instars. This is most likely caused by the immune system or incompatibility of the sexual line. Another well-described method for transinfecting *Wolbachia* is by injecting the egg (Hughes and Rasgon, 2014), which might be a way to circumvent part of the immune system. This technique is often used for Drosophila species, which unlike other insects, lack a serosa membrane (Schmidt-Ott, 2000; Rafiqi *et al.*, 2008). The serosa, an extraembryonic epithelium that covers the yolk and embryo, has been shown to be involved in the immune response by expressing antimicrobial peptides (Jacobs *et al.*, 2022). Collembola also lack a serosa, which might be an argument for transinfection through egg injection, although recent work suggests that at least in the springtail *Orchesella cincta* eggs can upregulate immune genes upon infection that are innately without a serosa (Cheng *et al.*, 2023). Thus, whether injections in eggs will lead to a less severe immune response and be more successful than haemolymph transfusions between adult *F. candida* remains uncertain.

Another option, which was not explored within this thesis, is transinfection through cannibalism. This is a more realistic scenario as this is a widespread

phenomenon even among non-carnivorous insects (Richardson *et al.*, 2010). Experiments using this method with both *Drosophila* and isopods gave mixed results, with lack of transinfection in the flies, but successful transinfections in isopods (Le Clec'h, Chevalier, *et al.*, 2013; Faria, Paulo and Sucena, 2016). Despite the uncertain outcomes of these alternative methods of transinfection, this research could prove worthwhile, as creating a new stable infection would be of great value to be able to track the development of a newly emerging parthenogenetic line of *F. candida*. This would give the opportunity to examine the changes that occur during this transition in both its traits, its genome and its interaction with *w*Fol.

Is *F. candida* a single species?

The first characterization of the sexual population already raised the question if *F. candida* is a single species (Frati *et al.*, 2004). My results emphasize the need to revisit this question, as chapter 4 demonstrated that the genomes of the sexual and asexual *F. candida* lineages differ greatly and chapter 1 suggested reproductive isolation between them. Can these lineages still be considered a single species?

As a widely accepted principle, the biological species concept states that a species is a group of organisms that can interbreed with the result of viable and fertile offspring (Mayr, 1942). When adhering strictly to this definition we should conclude that what has until now been described as *F. candida* are in fact separate species. Following this line of thought even further this would mean every parthenogenetic *F. candida* is a species of its own, as they do not mate with another individual before reproducing. The large genetic diversity between parthenogenetic lineages found by Frati et al. (2004) does suggest some form of speciation between asexual *F. candida*, although differences between siblings are too small to define them as separate species on those grounds.

For bacterial species the same issues apply. Here, for the practical reason of establishing common frame of reference by defining species, an operational definition is used based on arbitrary genetic thresholds (Konstantinidis, Ramette and Tiedje, 2006; Rodriguez-R *et al.*, 2024). A method closely linked to the phylogenetic species concept, where a species is defined as the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent, and that is distinguishable from other such clusters by a unique set of heritable characteristics (Staley, 2006). Whether *F. candida* adheres to this definition is highly debatable. As could be seen in the phylogeny of supplementary figure 4 within chapter 4, whether the sexual and asexual lineages fall within the same smallest diagnosable cluster with unique heritable characteristics depends strongly on the genomic marker used. However, the limited number of lineages and the relatively small fraction of the whole genome

considered in this analysis call for a more comprehensive approach to obtain a more definitive answer.

If such an analysis were to indicate that the sexual and asexual lineages represent distinct species, this would most likely reflect a case of *Wolbachia*-driven speciation, a mechanism hypothesized decades ago, although at the time in particular for cytoplasmic incompatibility (CI) strains (Breeuwer and Werren, 1990; Werren, 1997). This theory was recently confirmed with compelling evidence within both *Drosophila paulistorum* related to CI and in the weevil *Pantomorus postfasciatus* related to parthenogenesis inducing (PI) strains of *Wolbachia* (Baião *et al.*, 2019; Elias-Costa *et al.*, 2019). Thus, although this thesis gives some more insights on the matter, further research will need to prove if this is a case of *Wolbachia*-driven speciation as initially suspected by Frati et al. (2004).

Genes involved in PI

Since the start of *Wolbachia*-genomics, scientists have been searching for the genes responsible for its reproductive manipulations. After many attempts, this led to the discovery of the toxin-antitoxin genes responsible for CI (Beckmann, 2014; Beckmann, Ronau and Hochstrasser, 2017), encouraging the search for genes connected to the other manipulations. Within our *Wolbachia* comparative genomic study, no genes could be identified that were exclusively present in all PI *Wolbachia* strains. However, this does not fully exclude a genetic basis underlying PI, as the underlying mechanism is hypothesized to be more complicated than for CI, as it most likely developed multiple times and potentially requires various steps that differ across host species and their sex-determination systems (Ma and Schwander, 2017; Verhulst, Pannebakker and Geuverink, 2023). Therefore, the current expectation is that there are multiple sets of genes that can lead to parthenogenesis in organisms with different sex-determination systems.

Our genomic analysis yielded some genes that could potentially be interesting as they were found exclusively or almost exclusively within PI strains. However, no gene was found in all *Wolbachia* PI strains, strengthening the assumption that there are multiple genes linked to PI. Due to time constraints, we did not further investigate these genes to confirm their involvement in PI. Recently, another group did a more selective search in two genomes of Wolbachia strains carried by parasitoid wasps (*w*Tpre and *w*Lcla), in which parthenogenesis is induced by a similar two-step mechanism, wherein both diploidization and female-specific differentiation are required (Fricke and Lindsey, 2024). Two genes located in a region of phage origin were identified as strong contenders for causing PI labelled as pifA and pifB, which based on the locus tags are part of the set found by us. This highlights the strength of comparative genomics in the search for genes behind the reproductive manipulations and demonstrates once more that phage WO plays a big role in these processes.

Moreover, this validates that our search was going into the right direction and underscores the need for functional validation of the other loci identified. For the genes in *w*FoI, targeted knockouts or RNA interference (RNAi) to silence their expression would be particularly informative. While RNAi knockdown via injection has proven effective in the red flour beetle (*Tribolium castaneum*) (Posnien *et al.*, 2009), no such studies—to our knowledge—have yet been performed in Collembola. RNAi experiments could thus elucidate the functional roles of the candidates identified in our comparative genomics analysis.

To conclude this section on PI genes, it is worth addressing a question frequently posed by the general public: What are the practical benefits of studying parthenogenesis induction? To which I would state: Parthenogenesis circumvents the need for specific mating conditions in Hexapoda, often a limiting factor in laboratory culture, and substantially increases per-individual offspring yield through asexual reproduction. A recent opinion piece and review have made these benefits more concrete, highlighting applications such as the more efficient mass-rearing of parthenogenesis-inducing *Wolbachia*-infected parasitoid wasps for biological control (Leung *et al.*, 2020; Verhulst, Pannebakker and Geuverink, 2023). Future efforts to functionally characterize PI effectors could thus unlock scalable biocontrol solutions with significant agricultural impact.

What genes are linked to the degraded sexual traits

As phenotypic degradation of sexual traits in *F. candida* was described in the first chapter of this thesis and the full genomes of the sexual and asexual lines were sequenced in the last, it would have been insightful to determine if genetic degradation linked to this decay could be discovered. At the phenotypic level the decay was reflected in size of the spermatophore, its olfactory attractiveness, and the amount of sperm within it. At the genetic level, decay can arise either by relatively fast selection against these traits when they have negative pleiotropic effects or by slower and more stochastic mutation accumulation. The variable rates of decay in the phenotypic traits between the different *F. candida* lineages examined suggested that mutation accumulation is the main evolutionary process underlying this deterioration.

Decay by mutation accumulation is expected to leave a distinct genetic signal in the ratio of synonymous to non-synonymous mutations (Pamilo, Nei and li, 1987; Lynch and Blanchard, 1998; Henry, Schwander and Crespi, 2012; Brandt *et al.*, 2019). Mutation accumulation proceeds by drift and neutral processes, with an expected constant ratio of synonymous and non-synonymous mutations, while genes under negative selection are expected to deteriorate fast and would be expected to have a higher proportion of non-synonymous mutations or would be lost from the genome entirely (Wolf *et al.*, 2009; Guéguen and Duret, 2018).

Initial analyses were done on the genomes of the three lineages studied for male trait decay to test these hypotheses on genetic decay. A set of genes linked to male reproduction and a set of oogenesis genes, to contrast them as they are predicted not to decay, were identified based on a combination of orthogroups of genes and the annotation by similarity to other genes with known functions. Unfortunately, relatively few organisms with spermatophores have been studied in detail, and genes associated with this process are therefore poorly described. This has led to an inadequate definition of male sexual trait genes. Therefore, we needed to conclude that this analysis on genetic male sexual trait decay cannot be done with the knowledge currently available.

Genome quality is improving rapidly, as shown by the very recent publication of a haplotype-resolved assembly of the *F. candida* genome (Zhan, 2026), but annotation is still lagging behind. This might change drastically as the large language models (LLMs) are combined with quantum computing to make better predictions on genes functions a large scale. For LLMs it has already been demonstrated to enhance the annotation of viral proteins (Flamholz, Biller and Kelly, 2024; Harrigan *et al.*, 2024). When this develops towards full genome annotations, these questions might become easier to solve, although lab experiments will still need to be executed to check if those *in silico* predictions were correct.

Is trait loss in *F. candida* compensated by *w*FoI?

As my thesis focussed on the research question whether a switch from sexual to asexual reproduction under influence of an endosymbiont results in compensated trait loss, it is worth reflecting on this point. To show that compensated trait loss has occurred under these circumstances, you would need several lines of evidence that jointly demonstrate (i) loss of a male- or sex-related trait, (ii) counterbalancing of that loss by the endosymbiont, and (iii) relaxation of selection on the underlying genes.

In our study system, the loss of male- or sexual traits has been demonstrated clearly within this thesis and before by the observation on the lack of males within the asexual populations (Fountain and Hopkin, 2005; Kampfraath *et al.*, 2020). However, that the asexual reproduction is caused by Wolbachia and relaxation of selection on the underlying genes could not be proven unequivocally. As described in previous chapters and paragraphs, pinpointing Wolbachia as the source of PI will need either to find the exact genes in Wolbachia causative for PI or the creation of a newly infected lineage. Furthermore, to demonstrate the relaxation of selection we need to be able to specify the genes related to the male sexual traits.

Even though it was not possible to prove that F. candida its loss of sex was compensated by *w*FoI, all finding indicate that this is still the most likely hypothesis

and none of them contradicted this option. Therefore, I expect that when the science advances and improved methodologies are developed to study this question this will result in a positive answer to this question.

The bigger evolutionary genomic picture

To conclude, I would like to elaborate on a few observations related to genome structure and evolution that are not linked to the main questions of this thesis but deserve some further considerations.

The first is about genome size and how this can be linked to evolutionary pressures, which was triggered by our observation that the genome sizes of Wolbachia species correlate with those of their host. As stated before in chapter 3, this is in contradiction with the genome reduction theory for endosymbionts, which describes a widely seen phenomena that symbionts genome size is reduced depending on how closely they are bound to their host (Mira, Ochman and Moran, 2001; Ochman and Davalos, 2006; Moran, McCutcheon and Nakabachi, 2008; McCutcheon and Moran, 2012).

An alternative explanation could therefore be that the genome size is linked to the effective population size. In situations where the effective population size is low there is limited possibility for purifying selection and this could allow genome size to increase (Lynch and Conery, 2003). For both *F. candida* and *w*Fol, reduced effective population size is expected: the springtail reproduces asexually, limiting Ne, while its endosymbiont experiences severe bottlenecks during vertical transmission with minimal or no opportunity for horizontal acquisition of new strains. As it seems that the evolutionary trajectory here becomes similar for both host and endosymbiont, could this mean that as most endosymbionts have shrinking genomes the evolutionary pressures on their hosts push their genomes towards streamlining instead of conservation. A similar pattern to endosymbiont genome optimalisation is hypothesized for viruses, where larger viral genomes are associated with broader host ranges, while smaller genomes tend to be more host-restricted (Edwards, Steward and Schvarcz, 2021). Studying genomes of both bacterial and viral domains and their host could shed more light on the question whether there is a universal evolutionary trajectory behind this.

Secondly, what is striking is the pattern that there are regions in the genome that are very dynamic and regions that seem more conserved, a pattern first described in plant pathogenic fungi genomes and known as the two-speed genome model (Dong, Raffaele and Kamoun, 2015; Torres *et al.*, 2020). Our discovery of signs of two-speed genomes in both Wolbachia as their springtail host, suggest this might be a far wider phenomenon. Larger domain overarching studies need to be set-up to answer these questions, but the means to analyse such datasets in dept were beyond reach until recently. As for some time, we have been at a point that we are creating data faster than what we can analyse properly,

resulting in a surplus of data that has not been explored fully. If the promises made by AI and quantum computing companies hold true, this is about to turn-around and we will be in for a treat of unlocking more of the genome evolutionary mysteries.

REFERENCES



DKJR

References

Akman, L. *et al.* (2002) “Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia,” *Nature Genetics*, 32(3), pp. 402–407. Available at: https://doi.org/10.1038/ng986.

Altschul, S.F. *et al.* (1990) “Basic local alignment search tool,” *Journal of Molecular Biology*, 215(3), pp. 403–410. Available at: https://doi.org/10.1016/S0022-2836(05)80360-2.

Ament-Velásquez, S.L. *et al.* (2016) “Population genomics of sexual and asexual lineages in fissiparous ribbon worms (Lineus, Nemertea): hybridization, polyploidy and the Meselson effect,” *Molecular Ecology*, 25, pp. 3356–3369. Available at: https://doi.org/10.1111/mec.13717.

Archibald, J.M. (2015) “Endosymbiosis and eukaryotic cell evolution,” *Current Biology*, 25(19), pp. R911–R921. Available at: https://doi.org/10.1016/j.cub.2015.07.055.

Arndt, D. *et al.* (2016) “PHASTER: a better, faster version of the PHAST phage search tool,” *Nucleic acids research*, 44(W1), pp. W16–W21. Available at: https://doi.org/10.1093/nar/gkw387.

Bachtrog, D. (2013) “Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration,” *Nature Reviews Genetics*, 14(2), pp. 113–124. Available at: https://doi.org/10.1038/nrg3366.

Baião, G.C. *et al.* (2019) “The effect of Wolbachia on gene expression in Drosophila paulistorum and its implications for symbiont-induced host speciation,” *BMC Genomics*, 20(1), p. 465. Available at: https://doi.org/10.1186/s12864-019-5816-9.

Bao, W., Kojima, K.K. and Kohany, O. (2015) “Repbase Update, a database of repetitive elements in eukaryotic genomes,” *Mobile DNA*, 6(1), pp. 4–9. Available at: https://doi.org/10.1186/s13100-015-0041-9.

Barrick, J.E. and Lenski, R.E. (2013) “Genome dynamics during experimental evolution,” *Nature Reviews Genetics*, 14(12), pp. 827–839. Available at: https://doi.org/10.1038/nrg3564.

De Bary, A. (1879) *Die Erscheinung der Symbiose: Vortrag, gehalten auf der Versammlung deutscher Naturforscher und Aerzte zu Cassel*. Strassburg: Karl J. Trübner. Available at: https://archive.org/details/DieErscheinungDerSymbiose.

Beckmann, J.F. (2014) Molecular mechanism of Wolbachia induced cytoplasmic incompatibility. PhD thesis. University of Minnesota.

Beckmann, J.F., Ronau, J.A. and Hochstrasser, M. (2017) “A Wolbachia deubiquitylating enzyme induces cytoplasmic incompatibility,” Nature Microbiology, 2, article 17007. Available at: https://doi.org/10.1038/nmicrobiol.2017.7.

Bennett, G.M. and Moran, N.A. (2013) “Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect,” *Genome Biology and Evolution*, 5(9), pp. 1675–1688. Available at: https://doi.org/10.1093/gbe/evt118.

Bernt, M. *et al.* (2013) “MITOS: Improved de novo metazoan mitochondrial genome annotation,” *Molecular Phylogenetics and Evolution*, 69(2), pp. 313–319. Available at: https://doi.org/10.1016/j.ympev.2012.08.023.

Besemer, J. and Borodovsky, M. (2005) “GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses,” *Nucleic Acids Research*, 33(Web Server issue), pp. 451–454. Available at: https://doi.org/10.1093/nar/gki487.

Blanc-Mathieu, R. *et al.* (2017) *Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes*, *PLoS Genetics*, 13(6), article e1006777. Available at: https://doi.org/10.1371/journal.pgen.1006777.

De Bodt, S., Maere, S. and Van de Peer, Y. (2005) “Genome duplication and the origin of angiosperms,” *Trends in Ecology & Evolution*, 20(11), pp. 591–597. Available at: https://doi.org/10.1016/J.TREE.2005.07.008.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) “Trimmomatic: A flexible trimmer for Illumina sequence data,” *Bioinformatics*, 30(15), pp. 2114–2120. Available at: https://doi.org/10.1093/bioinformatics/btu170.

Bordenstein, Sarah R. and Bordenstein, Seth R. (2011) “Temperature affects the tripartite interactions between bacteriophage WO, Wolbachia, and cytoplasmic incompatibility,” *PLoS ONE*, 6(12), p. e29106. Available at: https://doi.org/10.1371/journal.pone.0029106.

Bordenstein, Sarah R. and Bordenstein, Seth R. (2016) “Eukaryotic association module in phage WO genomes from Wolbachia,” *Nature Communications*, 7, article 13155. Available at: https://doi.org/10.1038/ncomms13155.

Bordenstein, Sarah R. and Bordenstein, Seth R. (2022) *Widespread phages of endosymbionts: Phage WO genomics and the proposed taxonomic classification of Symbioviridae*, *PLOS Genetics*, 18(6), article e1010227. Available at: https://doi.org/10.1371/journal.pgen.1010227.

Bordenstein, S.R. *et al.* (2006) “The tripartite associations between bacteriophage, Wolbachia, and arthropods,” *PLoS Pathogens*, 2(5), pp. 384–393. Available at: https://doi.org/10.1371/journal.ppat.0020043.

Brandt, A. *et al.* (2019) “No signal of deleterious mutation accumulation in conserved gene sequences of extant asexual hexapods,” *Scientific Reports*, 9(1), p. 5338. Available at: https://doi.org/10.1038/s41598-019-41821-x.

Breeuwer, J.A.J. and Werren, J.H. (1990) “Microorganisms associated with chromosome destruction and reproductive isolation between two insect species,” *Nature*, 346(6284), pp. 558–560. Available at: https://doi.org/10.1038/346558a0.

Brown, A.M.V. *et al.* (2016) “Genomic evidence for plant-parasitic nematodes as the earliest Wolbachia hosts,” *Scientific Reports*, 6, article 34955. Available at: https://doi.org/10.1038/srep34955.

Buono, L. and Tully, T. (2011) Etude préliminaire des modes de reproduction chez Folsomia candida. Master 1 EBE report. Université Pierre et Marie Curie.

Burdina, E. V. *et al.* (2021) “Unique Wolbachia strain wMelPlus increases heat stress resistance in Drosophila melanogaster,” *Archives of Insect Biochemistry and Physiology*, 106(4), article e21776. Available at: https://doi.org/10.1002/arch.21776.

Burns, R.G. and Slater, H. (1982) “Experimental Microbial Ecology,” *Experimental Microbial Ecology*. Oxford: Blackwell Scientific Publications, pp. 66–83.

Campbell, A.M. (1992) “Chromosomal insertion sites for phages and plasmids,” *Journal of Bacteriology*, 174(23), pp. 7495–7499. Available at: https://doi.org/10.1128/jb.174.23.7495-7499.1992.

Cao, Z. *et al.* (2013) “The genome of Mesobuthus martensii reveals a unique adaptation model of arthropods,” *Nature Communications*, 4, pp. 1–10. Available at: https://doi.org/10.1038/ncomms3602.

Cerveau, N. *et al.* (2011) “Short- and long-term evolutionary dynamics of bacterial insertion sequences: Insights from Wolbachia endosymbionts,” *Genome Biology and Evolution*, 3(1), pp. 1175–1186. Available at: https://doi.org/10.1093/gbe/evr096.

Chafee, M.E. *et al.* (2011) “Decoupling of host-symbiont-phage coadaptations following transfer between insect species,” *Genetics*, 187(1), pp. 203–215. Available at: https://doi.org/10.1534/genetics.110.120675.

Chaisson, M.J. and Tesler, G. (2012) “Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory,” *BMC Bioinformatics*, 13(1), p. 238. Available at: https://doi.org/10.1186/1471-2105-13-238.

Cheng, S. *et al.* (2023) “Immune competence in eggs of the springtail *Orchesella cincta*,” in Cheng, S., *Innate immunity, developmental speed and their trade-offs in two hexapod models*. PhD thesis. Leiden University, chapter 2, pp. 31–41. Available at: https://hdl.handle.net/1887/3665319.

Chevreux, B., Wetter, T. and Suhai, S. (1999) “Genome sequence assembly using trace signals and additional sequence information,” in Computer science and biology: Proceedings of the German Conference on Bioinformatics (GCB ’99), pp. 45–56.

Chin, C.S. *et al.* (2013) “Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data,” *Nature Methods*, 10(6), pp. 563–569. Available at: https://doi.org/10.1038/nmeth.2474.

Chin, C.S. *et al.* (2016) “Phased diploid genome assembly with single-molecule real-time sequencing,” *Nature Methods*, 13(12), pp. 1050–1054. Available at: https://doi.org/10.1038/nmeth.4035.

Le Clec’h, W. *et al.* (2012) “High Virulence of Wolbachia after Host Switching: When Autophagy Hurts,” *PLoS Pathogens*, 8(8), p. e1002844. Available at: https://doi.org/10.1371/journal.ppat.1002844.

Le Clec'h, W., Chevalier, F.D., *et al.* (2013) "Cannibalism and Predation as Paths for Horizontal Passage of Wolbachia between Terrestrial Isopods," *PLoS ONE*, 8(4), p. e60232. Available at: https://doi.org/10.1371/journal.pone.0060232.

Le Clec'h, W., Raimond, M., *et al.* (2013) "Horizontal transfers of feminizing versus non-feminizing Wolbachia strains: From harmless passengers to pathogens," *Environmental Microbiology*, 15(11), pp. 2922–2936. Available at: https://doi.org/10.1111/1462-2920.12172.

Colbourne, J.K. *et al.* (2011) "The ecoresponsive genome of Daphnia pulex.," *Science (New York, N.Y.)*, 331(6017), pp. 555–61. Available at: https://doi.org/10.1126/science.1197761.

Comandatore, F. *et al.* (2015) "Supergroup C Wolbachia, mutualist symbionts of filarial nematodes, have a distinct genome structure," *Open Biology*, 5(12). Available at: https://doi.org/10.1098/rsob.150099.

Cooper, V.S. and Lenski, R.E. (2000) "The population genetics of ecological specialization in evolving Escherichia coli populations," *Nature*, 407(6805), pp. 736–739. Available at: https://doi.org/10.1038/35037572.

Cordaux, R., Bouchon, D. and Grève, P. (2011) "The impact of endosymbionts on the evolution of host sex-determination mechanisms," *Trends in Genetics*, 27(8), pp. 332–341. Available at: https://doi.org/10.1016/j.tig.2011.05.002.

Cotton, J.A. *et al.* (2016) "The genome of Onchocerca volvulus, agent of river blindness," *Nature Microbiology*, 2(November 2016), pp. 1–12. Available at: https://doi.org/10.1038/nmicrobiol.2016.216.

Czarnetzki, A.B. and Tebbe, C.C. (2004) "Detection and phylogenetic analysis of Wolbachia in Collembola," *Environmental Microbiology*, 6(1), pp. 35–44. Available at: https://doi.org/10.1046/j.1462-2920.2003.00537.x.

Dale, C. and Moran, N.A. (2006) "Molecular Interactions between Bacterial Symbionts and Their Hosts," *Cell*, 126(3), pp. 453–465. Available at: https://doi.org/10.1016/j.cell.2006.07.014.

Dallai, R. *et al.* (2014) "Giant spermatozoa and a huge spermatheca: A case of coevolution of male and female reproductive organs in the ground louse Zorotypus impolitus (Insecta, Zoraptera)," *Arthropod Structure and Development*, 43(2), pp. 135–151. Available at: https://doi.org/10.1016/j.asd.2013.10.002.

Dallai, R., Zizzari, Z. V. and Fanciulli, P.P. (2009) "Different sperm number in the spermatophores of Orchesella villosa (Geoffroy) (Entomobryidae) and Allacma fusca (L.) (Sminthuridae)," *Arthropod Structure and Development*, 38(3), pp. 227–234. Available at: https://doi.org/10.1016/j.asd.2008.10.002.

Dallai, R., Zizzari, Z.V. and Fanciulli, P.P. (2008) "Fine structure of the spermatheca and of the accessory glands in Orchesella villosa (Collembola, Hexapoda)," *Journal of Morphology*, 269(4), pp. 464–478. Available at: https://doi.org/10.1002/jmor.10595.

Darby, A.C. *et al.* (2012) "Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis," *Genome Research*, 22(12), pp. 2467–2477. Available at: https://doi.org/10.1101/gr.138420.112.

Darling, A.C.E. et al. (2004) "Mauve: multiple alignment of conserved genomic sequence with rearrangements," Genome Research, 14(7), pp. 1394–1403. Available at: https://doi.org/10.1101/gr.2289704.

Dedeine, F. *et al.* (2001) "Removing symbiotic Wolbachia bacteria specifically inhibits oogenesis in a parasitic wasp," *Proceedings of the National Academy of Sciences*, 98(11), pp. 6247–6252. Available at: https://doi.org/10.1073/pnas.101304298.

Depristo, M.A. *et al.* (2011) "A framework for variation discovery and genotyping using next-generation DNA sequencing data," *Nature Genetics*, 43(5), pp. 491–501. Available at: https://doi.org/10.1038/ng.806.

Dong, S., Raffaele, S. and Kamoun, S. (2015) "The two-speed genomes of filamentous pathogens: Waltz with plants," *Current Opinion in Genetics and Development*, 35, pp. 57–65. Available at: https://doi.org/10.1016/j.gde.2015.09.001.

Douglas, A.E. (1998) "Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera," Annual Review of Entomology, 43, pp. 17–37. Available at: https://doi.org/10.1146/annurev.ento.43.1.17.

Duplouy, A. *et al.* (2013) “Draft genome sequence of the male-killing Wolbachia strain wBol1 reveals recent horizontal gene transfers from diverse sources,” *BMC Genomics*, 14(1), p. 20. Available at: https://doi.org/10.1186/1471-2164-14-20.

Dy, R.L. *et al.* (2014) “A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism,” *Nucleic Acids Research*, 42(7), pp. 4590–4605. Available at: https://doi.org/10.1093/nar/gkt1419.

Edgar, R.C. (2004) “MUSCLE: multiple sequence alignment with high accuracy and high throughput.,” *Nucleic acids research*, 32(5), pp. 1792–7. Available at: https://doi.org/10.1093/nar/gkh340.

Edwards, K.F., Steward, G.F. and Schvarcz, C.R. (2021) “Making sense of virus size and the tradeoffs shaping viral fitness,” *Ecology Letters*, 24(2), pp. 363–373. Available at: https://doi.org/10.1111/ele.13630.

Elias-Costa, A.J. *et al.* (2019) “Game of clones: Is Wolbachia inducing speciation in a weevil with a mixed reproductive mode?,” *Molecular Phylogenetics and Evolution*, 133(December 2018), pp. 42–53. Available at: https://doi.org/10.1016/j.ympev.2018.12.027.

Ellegaard, K.M. *et al.* (2013) “Comparative Genomics of Wolbachia and the Bacterial Species Concept,” *PLoS Genetics*, 9(4), p. e1003381. Available at: https://doi.org/10.1371/journal.pgen.1003381.

Ellers, J. *et al.* (2012) “Ecological interactions drive evolutionary loss of traits,” *Ecology Letters*, 15(10), pp. 1071–1082. Available at: https://doi.org/10.1111/j.1461-0248.2012.01830.x.

Emanuel, B.S. and Shaikh, T.H. (2001) “Segmental duplications: an ‘expanding’ role in genomic instability and disease,” *Nature Reviews Genetics*, 2(10), pp. 791–800. Available at: https://doi.org/10.1038/35093500.

Emms, D.M. and Kelly, S. (2015) “OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy,” *Genome Biology*, 16(1), p. 157. Available at: https://doi.org/10.1186/s13059-015-0721-2.

English, A.C. *et al.* (2012) “Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology,” *PLoS ONE*, 7(11), pp. 1–12. Available at: https://doi.org/10.1371/journal.pone.0047768.

Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002) “An efficient algorithm for large-scale detection of protein families,” *Nucleic Acids Research*, 30(7), pp. 1575–1584. Available at: https://doi.org/10.1093/nar/30.7.1575.

Faddeeva-Vakhrusheva, A. *et al.* (2016) “Gene family evolution reflects adaptation to soil environmental stressors in the genome of the collembolan orchesella cincta,” *Genome Biology and Evolution*, 8(7), pp. 2106–2117. Available at: https://doi.org/10.1093/gbe/evw134.

Faddeeva-Vakhrusheva, A. *et al.* (2017) “Coping with living in the soil: The genome of the parthenogenetic springtail Folsomia candida,” *BMC Genomics*, 18(1), pp. 1–14. Available at: https://doi.org/10.1186/s12864-017-3852-x.

Faria, R. and Navarro, A. (2010) “Chromosomal speciation revisited: Rearranging theory with pieces of evidence,” *Trends in Ecology and Evolution*, 25(11), pp. 660–669. Available at: https://doi.org/10.1016/j.tree.2010.07.008.

Faria, V.G., Paulo, T.F. and Sucena, É. (2016) “Testing cannibalism as a mechanism for horizontal transmission of Wolbachia in Drosophila,” *Symbiosis*, 68(1–3), pp. 79–85. Available at: https://doi.org/10.1007/s13199-015-0354-y.

Ferguson, L.F., Ross, P.A. and van Heerwaarden, B. (2024) “Wolbachia infection negatively impacts Drosophila simulans heat tolerance in a strain- and trait-specific manner,” *Environmental Microbiology*, 26(4), pp. 1–22. Available at: https://doi.org/10.1111/1462-2920.16609.

Finn, R.D. *et al.* (2014) “Pfam: The protein families database,” *Nucleic Acids Research*, 42(D1), pp. 222–230. Available at: https://doi.org/10.1093/nar/gkt1223.

Fisher, R.M. *et al.* (2017) “The evolution of host-symbiont dependence,” *Nature Communications*, 8(May), pp. 1–8. Available at: https://doi.org/10.1038/ncomms15973.

Flamholz, Z.N., Biller, S.J. and Kelly, L. (2024) “Large language models improve annotation of prokaryotic viral proteins,” *Nature Microbiology*, 9(2), pp. 537–549. Available at: https://doi.org/10.1038/s41564-023-01584-8.

Flot, J.F. *et al.* (2013) “Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga,” *Nature*, 500(7463), pp. 453–457. Available at: https://doi.org/10.1038/nature12326.

Fong, D.W., Kane, T.C. and Culver, D.C. (1995) “Vestigialization and loss of nonfunctional characters,” *Annual Review of Ecology and Systematics*, 26, pp. 249–268. Available at: https://doi.org/10.1146/annurev.es.26.110195.001341.

Foster, J. *et al.* (2005) “The Wolbachia genome of Brugia malayi: Endosymbiont evolution within a human pathogenic nematode,” *PLoS Biology*, 3(4), pp. 0599–0614. Available at: https://doi.org/10.1371/journal.pbio.0030121.

Fountain, M.T. and Hopkin, S.P. (2005) “*FOLSOMIA CANDIDA* (COLLEMBOLA): A ‘Standard’ Soil Arthropod,” *Annual Review of Entomology*, 50(1), pp. 201–222. Available at: https://doi.org/10.1146/annurev.ento.50.071803.130331.

Francis, W.R., Vargas, S. and Wörheide, G. (2025) “Genomic changes are varied across congeneric species pairs of animals,” *Peer Community Journal*, 5, p. 0. Available at: https://doi.org/10.24072/pcjournal.563.

Frati, F. *et al.* (2004) “High levels of genetic differentiation between Wolbachia-infected and non-infected populations of Folsomia candida (Collembola, Isotomidae),” *Pedobiologia*, 48(5–6), pp. 461–468. Available at: https://doi.org/10.1016/j.pedobi.2004.04.004.

Fricke, L.C. and Lindsey, A.R.I. (2024) “Identification of Parthenogenesis-Inducing Effector Proteins in Wolbachia,” *Genome Biology and Evolution*, 16(4), pp. 1–14. Available at: https://doi.org/10.1093/gbe/evae036.

Gerardo, N.M., Hoang, K.L. and Stoy, K.S. (2020) “Evolution of animal immunity in the light of beneficial symbioses,” Philosophical Transactions of the Royal Society B: Biological Sciences, 375(1808), article 20190601. Available at: https://doi.org/10.1098/rstb.2019.0601.

Gerth, M. *et al.* (2014) “Phylogenomic analyses uncover origin and spread of the Wolbachia pandemic,” *Nature Communications*, 5(iDiv), p. 5117. Available at: https://doi.org/10.1038/ncomms6117.

Gilbert, C., Peccoud, J. and Cordaux, R. (2021) “Transposable Elements and the Evolution of Insects,” *Annual Review of Entomology*, 66, pp. 355–372. Available at: https://doi.org/10.1146/annurev-ento-070720-074650.

Gill, A.C., Darby, A.C. and Makepeace, B.L. (2014) “Iron Necessity: The Secret of Wolbachia’s Success?,” *PLoS Neglected Tropical Diseases*, 8(10), p. e3224. Available at: https://doi.org/10.1371/journal.pntd.0003224.

Giordano, R. *et al.* (2010) “Effect of a High Dose of Three Antibiotics on the Reproduction of a Parthenogenetic Strain of <I>Folsomia candida</I> (Isotomidae: Collembola),” *Environmental Entomology*, 39(4), pp. 1170–1177. Available at: https://doi.org/10.1603/EN10027.

Giorgini, M. (2001) “Induction of males in thelytokous populations of Encarsia meritoria and Encarsia protransvena: A systematic tool,” *BioControl*, 46(4), pp. 427–438. Available at: https://doi.org/10.1023/A:1014181431482.

Gols, R., Ernsting, G. and Van Straalen, N.M. (2004) “Paternity analysis in a hexapod (Orchesella cincta; Collembola) with indirect sperm transfer,” *Journal of Insect Behavior*, 17(3), pp. 317–328. Available at: https://doi.org/10.1023/B:JOIR.0000031533.32859.ba.

Goto, H.E. (1960) “Facultative parthenogenesis in Collembola (Insecta),” Nature, 188, pp. 958–959. Available at: https://doi.org/10.1038/188958a0.

Gottlieb, Y. *et al.* (2002) “Diploidy restoration in Wolbachia-infected Muscidifurax uniraptor (Hymenoptera: Pteromalidae),” *Journal of Invertebrate Pathology*, 81(3), pp. 166–174. Available at: https://doi.org/10.1016/S0022-2011(02)00149-0.

Gottlieb, Y. and Zchori-Fein, E. (2001) “Irreversible thelytokous reproduction in Muscidifurax uniraptor,” *Entomologia Experimentalis et Applicata*, 100(3), pp. 271–278. Available at: https://doi.org/10.1023/A:1019298825049.

Groth, A.C. and Calos, M.P. (2004) “Phage integrases: Biology and applications,” *Journal of Molecular Biology*, 335(3), pp. 667–678. Available at: https://doi.org/10.1016/j.jmb.2003.09.082.

Gu, X. *et al.* (2022) “A wMel Wolbachia variant in Aedes aegypti from field-collected Drosophila melanogaster with increased phenotypic stability under heat stress,” *Environmental Microbiology*, 24(4), pp. 2119–2135. Available at: https://doi.org/10.1111/1462-2920.15966.

Guéguen, L. and Duret, L. (2018) “Unbiased estimate of synonymous and nonsynonymous substitution rates with nonstationary base composition,” *Molecular Biology and Evolution*, 35(3), pp. 734–742. Available at: https://doi.org/10.1093/molbev/msx308.

Hall, A.R. and Colegrave, N. (2008) “Decay of unused characters by selection and drift,” *Journal of Evolutionary Biology*, 21(2), pp. 610–617. Available at: https://doi.org/10.1111/j.1420-9101.2007.01473.x.

Hall, B.G. (2013) “Building phylogenetic trees from molecular data with MEGA,” *Molecular Biology and Evolution*, 30(5), pp. 1229–1235. Available at: https://doi.org/10.1093/molbev/mst012.

Harrigan, W.L. *et al.* (2024) “Improvements in viral gene annotation using large language models and soft alignments,” *BMC Bioinformatics*, 25(1), pp. 1–19. Available at: https://doi.org/10.1186/s12859-024-05779-6.

Havird, J.C., Hall, M.D. and Dowling, D.K. (2015) “The evolution of sex: a new hypothesis based on mitochondrial mutational erosion,” BioEssays, 37(9), pp. 951–958. Available at: https://doi.org/10.1002/bies.201500057.

Heasley, L.R. and Argueso, J.L. (2022) “Bursts of Genomic Instability Potentiate Phenotypic and Genomic Diversification in Saccharomyces cerevisiae,” *Frontiers in Genetics*, 13(June), pp. 1–12. Available at: https://doi.org/10.3389/fgene.2022.912851.

Heethoff, M. *et al.* (2007) “High genetic divergences indicate ancient separation of parthenogenetic lineages of the oribatid mite Platynothrus peltifer (Acari, Oribatida),” *Journal of Evolutionary Biology*, 20(1), pp. 392–402. Available at: https://doi.org/10.1111/j.1420-9101.2006.01183.x.

Henry, L., Schwander, T. and Crespi, B.J. (2012) “Deleterious mutation accumulation in asexual timema stick insects,” *Molecular Biology and Evolution*, 29(1), pp. 401–408. Available at: https://doi.org/10.1093/molbev/msr237.

Hilgenboecker, K. *et al.* (2008) “How many species are infected with Wolbachia? - A statistical analysis of current data,” *FEMS Microbiology Letters*, 281(2), pp. 215–220. Available at: https://doi.org/10.1111/j.1574-6968.2008.01110.x.

Hoff, K.J. *et al.* (2016) “BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS,” *Bioinformatics*, 32(5), pp. 767–769. Available at: https://doi.org/10.1093/bioinformatics/btv661.

Hollister, J.D. *et al.* (2015) “Recurrent loss of sex is associated with accumulation of deleterious mutations in oenothera,” *Molecular Biology and Evolution*, 32(4), pp. 896–905. Available at: https://doi.org/10.1093/molbev/msu345.

Holt, C. and Yandell, M. (2011) “MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects,” BMC Bioinformatics, 12, article 491. Available at: https://doi.org/10.1186/1471-2105-12-491.

Hubley, R. *et al.* (2016) “The Dfam database of repetitive DNA families,” *Nucleic Acids Research*, 44(D1), pp. D81–D89. Available at: https://doi.org/10.1093/nar/gkv1272.

Hughes, G.L. and Rasgon, J.L. (2014) “Transinfection: A method to investigate Wolbachia-host interactions and control arthropod-borne disease,” *Insect Molecular Biology*, 23(2), pp. 141–151. Available at: https://doi.org/10.1111/imb.12066.

Hughes, J.F. *et al.* (2005) “Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee,” *Nature*, 437(7055), pp. 101–104. Available at: https://doi.org/10.1038/nature04101.

Huigens, M.E. *et al.* (2004) “Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing Wolbachia in Trichogramma wasps,” *Proceedings of the Royal Society B: Biological Sciences*, 271(1538), pp. 509–515. Available at: https://doi.org/10.1098/rspb.2003.2640.

Hyatt, D. *et al.* (2010) “Prodigal: Prokaryotic gene recognition and translation initiation site identification,” *BMC Bioinformatics*, 11. Available at: https://doi.org/10.1186/1471-2105-11-119.

National Center for Biotechnology Information (n.d.) VecScreen: screen a sequence for vector contamination. Available at: https://www.ncbi.nlm.nih.gov/tools/vecscreen/

Iturbe-Ormaetxe, I. *et al.* (2005) “Distribution, expression, and motif variability of ankyrin domain genes in Wolbachia pipientis,” *Journal of Bacteriology*, 187(15), pp. 5136–5145. Available at: https://doi.org/10.1128/JB.187.15.5136-5145.2005.

Jacobs, C.G.C. *et al.* (2022) “Immune function of the serosa in hemimetabolous insect eggs,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1865). Available at: https://doi.org/10.1098/rstb.2021.0266.

Jaenike, J. (2007) “Spontaneous emergence of a new Wolbachia phenotype,” *Evolution*, 61(9), pp. 2244–2252. Available at: https://doi.org/10.1111/j.1558-5646.2007.00180.x.
Jalinsky, J., Logsdon, J.M. and 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), pp. 1050–1059. Available at: https://doi.org/10.1111/jeb.13632.
Jaron, K.S. *et al.* (2020) “Genomic Features of Parthenogenetic Animals,” *Journal of Heredity*, pp. 1–15. Available at: https://doi.org/10.1093/jhered/esaa031.
Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) “The rapid generation of mutation data matrices from protein sequences,” *Bioinformatics*, 8(3), pp. 275–282. Available at: https://doi.org/10.1093/bioinformatics/8.3.275.
Jones, P. *et al.* (2014) “InterProScan 5: Genome-scale protein function classification,” *Bioinformatics*, 30(9), pp. 1236–1240. Available at: https://doi.org/10.1093/bioinformatics/btu031.
Jonkheer, E.M. *et al.* (2025) “Exploring intra- and intergenomic variation in haplotype-resolved pangenomes,” *Plant Biotechnology Journal*, 23(3), pp. 874–886. Available at: https://doi.org/10.1111/pbi.14545.
Kambris, Z. et al. (2009) “Immune activation by life-shortening Wolbachia and reduced filarial competence in mosquitoes,” Science, 326(5949), pp. 134–136. Available at: https://doi.org/10.1126/science.1177531.
Kampfraath, A.A. *et al.* (2019) “Genome expansion of an obligate parthenogenesis-associated Wolbachia poses an exception to the symbiont reduction model,” *BMC Genomics*, 20(106), pp. 1–14. Available at: https://doi.org/10.1186/s12864-019-5492-9.
Kampfraath, A.A. et al. (2020) “Male sexual trait decay in two asexual springtail populations follows neutral mutation accumulation theory,” Evolutionary Biology, 47(4), pp. 285–292. Available at: https://doi.org/10.1007/s11692-020-09511-z.
Keeling, P.J. and McCutcheon, J.P. (2017) “Endosymbiosis: The feeling is not mutual,” *Journal of Theoretical Biology*, 434, pp. 75–79. Available at: https://doi.org/10.1016/j.jtbi.2017.06.008.
Kent, B.N. *et al.* (2011) “Evolutionary genomics of a temperate bacteriophage in an obligate intracellular bacteria (Wolbachia),” *PLoS ONE*, 6(9), p. e24984. Available at: https://doi.org/10.1371/journal.pone.0024984.
Kent, B.N. and Bordenstein, S.R. (2010) “Phage WO of Wolbachia: Lambda of the endosymbiont world,” *Trends in Microbiology*, 18(4), pp. 173–181. Available at: https://doi.org/10.1016/j.tim.2009.12.011.
Kimura, M. (1983) *The neutral theory of molecular evolution*. Cambridge University Press.
Klasson, L. *et al.* (2008) “Genome evolution of Wolbachia strain wPip from the Culex pipiens group,” *Molecular Biology and Evolution*, 25(9), pp. 1877–1887. Available at: https://doi.org/10.1093/molbev/msn133.
Klasson, L. *et al.* (2009) “The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans,” *Proceedings of the National Academy of Sciences*, 106(14), pp. 5725–5730. Available at: https://doi.org/10.1073/pnas.0810753106.
Koivisto, R.K.K. and Braig, H.R. (2003) “Microorganisms and parthenogenesis,” *Biological Journal of the Linnean Society*, 79(1), pp. 43–58. Available at: https://doi.org/10.1046/j.1095-8312.2003.00185.x.
Konstantinidis, K.T., Ramette, A. and Tiedje, J.M. (2006) “The bacterial species definition in the genomic era,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1475), pp. 1929–1940. Available at: https://doi.org/10.1098/rstb.2006.1920.
van der Kooi, C.J. and Schwander, T. (2014) “On the fate of sexual traits under asexuality,” *Biological Reviews*, 89(4), pp. 805–819. Available at: https://doi.org/10.1111/brv.12078.
Koren, S. *et al.* (2017) “Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation,” *Genome Research*, 27(5), pp. 722–736. Available at: https://doi.org/10.1101/gr.215087.116.
Kraaijeveld, K. *et al.* (2009) “Effects of parthenogenesis and geographic isolation on female sexual traits in a parasitoid WASP,” *Evolution*, 63(12), pp. 3085–3096. Available at: https://doi.org/10.1111/j.1558-5646.2009.00798.x.

Kraaijeveld, K. *et al.* (2012) “Transposon proliferation in an asexual parasitoid,” *Molecular Ecology*, 21(16), pp. 3898–3906. Available at: https://doi.org/10.1111/j.1365-294X.2012.5582.x.
Kraaijeveld, K. *et al.* (2016) “Decay of sexual trait genes in an asexual parasitoidwasp,” *Genome Biology and Evolution*, 8(12), pp. 3685–3695. Available at: https://doi.org/10.1093/gbe/evw273.
Krogh, P.H. *et al.* (2008) “Toxicity testing with the collembolans Folsomia fimetaria and Folsomia candida and the results of a ringtest,” *Ecology*, 23(August), pp. 1–44. Available at: https://doi.org/10.1007/BF02703726.
Krzywinski, M. *et al.* (2009) “Circos: an information aesthetic for comparative genomics.,” *Genome research*, 19(9), pp. 1639–45. Available at: https://doi.org/10.1101/gr.092759.109.
Kumar, S., Stecher, G. and Tamura, K. (2016) “MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets,” *Molecular biology and evolution*, 33(7), pp. 1870–1874. Available at: https://doi.org/10.1093/molbev/msw054.
Laetsch, D.R. and Blaxter, M.L. (2017) “BlobTools: interrogation of genome assemblies,” F1000Research, 6, article 1287, version 1. Available at: https://doi.org/10.12688/f1000research.12232.1.
Lagesen, K. *et al.* (2007) “RNAmmer: Consistent and rapid annotation of ribosomal RNA genes,” *Nucleic Acids Research*, 35(9), pp. 3100–3108. Available at: https://doi.org/10.1093/nar/gkm160.
Lahti, D.C. *et al.* (2009) “Relaxed selection in the wild,” *Trends in Ecology and Evolution*, 24(9), pp. 487–496. Available at: https://doi.org/10.1016/j.tree.2009.03.010.
Leclercq, S. *et al.* (2016) “Birth of a W sex chromosome by horizontal transfer of *Wolbachia* bacterial symbiont genome,” *Proceedings of the National Academy of Sciences*, 113(52), pp. 15036–15041. Available at: https://doi.org/10.1073/pnas.1608979113.
Lefébure, T. *et al.* (2017) “Less effective selection leads to larger genomes,” *Genome Research*, 27(6), pp. 1016–1028. Available at: https://doi.org/10.1101/gr.212589.116.
Leung, K. et al. (2020) “Next-generation biological control: the need for integrating genetics and genomics,” Biological Reviews, 95(6), pp. 1838–1854. Available at: https://doi.org/10.1111/brv.12641.
Leung, W. *et al.* (2015) “ Drosophila Muller F Elements Maintain a Distinct Set of Genomic Properties Over 40 Million Years of Evolution ,” *G3: Genes|Genomes|Genetics*, 5(5), pp. 719–740. Available at: https://doi.org/10.1534/g3.114.015966.
Levasseur, A. and Pontarotti, P. (2011) “The role of duplications in the evolution of genomes highlights the need for evolutionary-based approaches in comparative genomics,” *Biology Direct*, 6(1), p. 11. Available at: https://doi.org/10.1186/1745-6150-6-11.
Li, H. and Durbin, R. (2009) “Fast and accurate short read alignment with Burrows-Wheeler transform,” *Bioinformatics*, 25(14), pp. 1754–1760. Available at: https://doi.org/10.1093/bioinformatics/btp324.
Lindsey, A.R.I. *et al.* (2016) “Comparative Genomics of a Parthenogenesis-Inducing *Wolbachia* Symbiont,” *G3: Genes|Genomes|Genetics*, 6(7), pp. 2113–2123. Available at: https://doi.org/10.1534/g3.116.028449.
Lindsey, A.R.I. *et al.* (2018) “Evolutionary Genetics of Cytoplasmic Incompatibility Genes cifA and cifB in Prophage WO of Wolbachia,” *Genome Biology and Evolution*, 10(2), pp. 434–451. Available at: https://doi.org/10.1093/gbe/evy012.
Lo, W.S., Huang, Y.Y. and Kuo, C.H. (2016) “Winding paths to simplicity: Genome evolution in facultative insect symbionts,” *FEMS Microbiology Reviews*, 40(6), pp. 855–874. Available at: https://doi.org/10.1093/femsre/fuw028.
Lovell, J.T. *et al.* (2017) “Mutation Accumulation in an Asexual Relative of Arabidopsis,” *PLoS Genetics*, 13(1), pp. 1–14. Available at: https://doi.org/10.1371/journal.pgen.1006550.
Lowe, T.M. and Eddy, S.R. (1997) “tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence,” Nucleic Acids Research, 25(5), pp. 955–964. Available at: https://doi.org/10.1093/nar/25.5.955.
Lu, M.H., Zhang, K.J. and Hong, X.Y. (2012) “Tripartite associations among bacteriophage WO, Wolbachia, and host affected by temperature and age in Tetranychus urticae,” *Experimental and Applied Acarology*, 58(3), pp. 207–220. Available at: https://doi.org/10.1007/s10493-012-9578-1.

Luan, Y.X. *et al.* (2023) “High-quality genomes reveal significant genetic divergence and cryptic speciation in the model organism Folsomia candida (collembola),” *Molecular Ecology Resources*, 23(1), pp. 273–293. Available at: https://doi.org/10.1111/1755-0998.13699.
Lynch, M. and Blanchard, J.L. (1998) “Deleterious mutation accumulation in organelle genomes,” Genetica, 102–103, pp. 29–39. Available at: https://doi.org/10.1023/A:1017022522486.
Lynch, M. and Conery, J.S. (2003) “The Origins of Genome Complexity,” *Science*, 302(5649), pp. 1401–1404. Available at: https://doi.org/10.1126/science.1089370.
Ma, W.J. *et al.* (2014) “Genetics of decayed sexual traits in a parasitoid wasp with endosymbiont-induced asexuality,” *Heredity*, 113(5), pp. 424–431. Available at: https://doi.org/10.1038/hdy.2014.43.
Ma, W.J. *et al.* (2015) “Diploid males support a two-step mechanism of endosymbiont-induced thelytoky in a parasitoid wasp Experimental evolution,” *BMC Evolutionary Biology*, 15(1), p. 84. Available at: https://doi.org/10.1186/s12862-015-0370-9.
Ma, W.J. and Schwander, T. (2017) “Patterns and mechanisms in instances of endosymbiont-induced parthenogenesis,” *Journal of Evolutionary Biology*, 30(5), pp. 868–888. Available at: https://doi.org/10.1111/jeb.13069.
Masui, S. *et al.* (2000) “Distribution and evolution of bacteriophage WO in Wolbachia, the endosymbiont causing sexual alterations in Arthropods,” *Journal of Molecular Evolution*, 51(5), pp. 491–497. Available at: https://doi.org/10.1007/s002390010112.
Mayr, E. (1942) *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.
McCutcheon, J.P. and Moran, N.A. (2012) “Extreme genome reduction in symbiotic bacteria,” *Nature Reviews Microbiology*, 10(1), pp. 13–26. Available at: https://doi.org/10.1038/nrmicro2670.
Merçot, H. and Poinsot, D. (2009) “Infection by Wolbachia: from passengers to residents,” *Comptes Rendus - Biologies*, 332(2–3), pp. 284–297. Available at: https://doi.org/10.1016/j.crvi.2008.09.010.
Mesbah-Uddin, M. *et al.* (2018) “Genome-wide mapping of large deletions and their population-genetic properties in dairy cattle,” *DNA Research*, 25(1), pp. 49–59. Available at: https://doi.org/10.1093/dnares/dsx037.
Metcalf, J.A. *et al.* (2014) “Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism,” *PeerJ*, 2, p. e529. Available at: https://doi.org/10.7717/peerj.529.
Mioduchowska, M. *et al.* (2023) “Playing Peekaboo with a Master Manipulator: Metagenetic Detection and Phylogenetic Analysis of Wolbachia Supergroups in Freshwater Invertebrates,” *International Journal of Molecular Sciences*, 24(11). Available at: https://doi.org/10.3390/ijms24119400.
Mira, A., Ochman, H. and Moran, N.A. (2001) “Deletional bias and the evolution of bacterial genomes,” *Trends in Genetics*, 17(10), pp. 589–596. Available at: https://doi.org/10.1016/S0168-9525(01)02447-7.
Moran, N. a and Mira, a (2001) “The process of genome shrinkage in the obligate symbiont Buchnera aphidicola.,” *Genome biology*, 2(12), p. RESEARCH0054. Available at: https://doi.org/10.1186/gb-2001-2-12-research0054.
Moran, N.A. (1996) “Accelerated evolution and Muller’s rachet in endosymbiotic bacteria.,” *Proceedings of the National Academy of Sciences*, 93(7), pp. 2873–2878. Available at: https://doi.org/10.1073/pnas.93.7.2873.
Moran, N.A. and Bennett, G.M. (2014) “The Tiniest Tiny Genomes,” *Annual Review of Microbiology*, 68(1), pp. 195–215. Available at: https://doi.org/10.1146/annurev-micro-091213-112901.
Moran, N.A., McCutcheon, J.P. and Nakabachi, A. (2008) “Genomics and Evolution of Heritable Bacterial Symbionts,” *Annual Review of Genetics*, 42(1), pp. 165–190. Available at: https://doi.org/10.1146/annurev.genet.41.110306.130119.
Moran, R.L. *et al.* (2023) “Selection-driven trait loss in independently evolved cavefish populations,” *Nature Communications*, 14(1), pp. 1–19. Available at: https://doi.org/10.1038/s41467-023-37909-8.
Muller, H.J. (1949) “The Darwinian and Modern Conceptions of Natural Selection,” *Proceedings of the American Philosophical Society*, 93(6), pp. 459–470.

Myers, E.W. et al. (2000) “A whole-genome assembly of Drosophila,” Science, 287(5461), pp. 2196–2204. Available at: https://doi.org/10.1126/science.287.5461.2196.
Newton, I.L.G. *et al.* (2016) “Comparative genomics of two closely related wolbachia with different reproductive effects on hosts,” *Genome Biology and Evolution*, 8(5), pp. 1526–1542. Available at: https://doi.org/10.1093/gbe/evw096.
Nikoh, N. *et al.* (2014) “Evolutionary origin of insect-Wolbachia nutritional mutualism,” *Proceedings of the National Academy of Sciences*, 111(28), pp. 10257–10262. Available at: https://doi.org/10.1073/pnas.1409284111.
Normark, B.B., Judson, O.P. and Moran, N.A. (2003) “Genomic signatures of ancient asexual lineages,” Biological Journal of the Linnean Society, 79(1), pp. 69–84. Available at: https://doi.org/10.1046/j.1095-8312.2003.00182.x.
Nowell, R.W. *et al.* (2018) *Comparative genomics of bdelloid rotifers: Insights from desiccating and nondesiccating species*, *PLoS Biology*. Available at: https://doi.org/10.1371/journal.pbio.2004830.
Nuzhdin, S.V. and Petrov, D.A. (2003) “Transposable elements in clonal lineages: lethal hangover from sex,” Biological Journal of the Linnean Society, 79(1), pp. 33–41. Available at: https://doi.org/10.1046/j.1095-8312.2003.00188.x.
Ochman, H. and Davalos, L.M. (2006) “The Nature and Dynamics of Bacterial Genomes,” *Science*, 311(March), pp. 1730–1733. Available at: https://doi.org/10.1126/science.1119966.
Olanratmanee, P. *et al.* (2021) “Novel supergroup u wolbachia in bat mites of Thailand,” *Southeast Asian Journal of Tropical Medicine and Public Health*, 52(1), pp. 48–55.
Oliver, J.H., Tanaka, K. and Sawada, M. (1973) “Cytogenetics of ticks (Acari: Ixodoidea) - 12. Chromosome and hybridization studies of bisexual and parthenogenetic Haemaphysalis longicornis races from Japan and Korea,” *Chromosoma*, 42(3), pp. 269–288. Available at: https://doi.org/10.1007/BF00284775.
Le Page, D.P. *et al.* (2017) “Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility,” *Nature*, 543(7644), pp. 243–247. Available at: https://doi.org/10.1038/nature21391.
Pamilo, P., Nei, M. and li, W.H. (1987) “Accumulation of mutations in sexual and asexual populations,” *Genetical Research*, 49(2), pp. 135–146. Available at: https://doi.org/10.1017/S0016672300026938.
Pamilo, P., Viljakainen, L. and Vihavainen, A. (2007) “Exceptionally high density of NUMTs in the honeybee genome,” Molecular Biology and Evolution, 24(6), pp. 1340–1346. Available at: https://doi.org/10.1093/molbev/msm055.
Pannebakker, B.A. *et al.* (2004) “Cytology of *Wolbachia* -induced parthenogenesis in *Leptopilina clavipes* (Hymenoptera: Figitidae),” *Genome*, 47(2), pp. 299–303. Available at: https://doi.org/10.1139/g03-137.
Parfrey, L.W., Lahr, D.J.G. and Katz, L.A. (2008) “The dynamic nature of eukaryotic genomes,” Molecular Biology and Evolution, 25(4), pp. 787–794. Available at: https://doi.org/10.1093/molbev/msn032.
Pati, A. *et al.* (2010) “GenePRIMP: A gene prediction improvement pipeline for prokaryotic genomes,” *Nature Methods*, 7(6), pp. 455–457. Available at: https://doi.org/10.1038/nmeth.1457.
Pearson, W.R. and Lipman, D.J. (1988) “Improved tools for biological sequence comparison.,” *Proceedings of the National Academy of Sciences*, 85(8), pp. 2444–2448. Available at: https://doi.org/10.1073/pnas.85.8.2444.
Van de Peer, Y., Maere, S. and Meyer, A. (2009) “The evolutionary significance of ancient genome duplications,” *Nature Reviews Genetics*, 10(10), pp. 725–732. Available at: https://doi.org/10.1038/nrg2600.
Petersen, M. *et al.* (2019) “Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects,” *BMC Evolutionary Biology*, 19(1), pp. 1–15. Available at: https://doi.org/10.1186/s12862-018-1324-9.
Pike, N. and Kingcombe, R. (2009) “Antibiotic treatment leads to the elimination of Wolbachia endosymbionts and sterility in the diplodiploid collembolan Folsomia candida,” *BMC Biology*, 7(54), p. 54. Available at: https://doi.org/10.1186/1741-7007-7-54.

Poggendorf, D. (1956) “Über rhythmische sexuelle Aktivität und ihre Beziehungen zu Häutung und Haarbildung bei arthropleonen Collembolen (Springschwanze),” *Naturwissenschaften*, 43(2), pp. 45–45.
Posnien, N. *et al.* (2009) “RNAi in the red flour beetle (Tribolium),” *Cold Spring Harbor Protocols*, 4(8), pp. 1–8. Available at: https://doi.org/10.1101/pdb.prot5256.
Pradhan, G. and Engsontia, P. (2023) “Diversity of the Antimicrobial Peptide Genes in Collembola,” *Insects*, 14(3), pp. 1–19. Available at: https://doi.org/10.3390/insects14030215.
Presgraves, D.C., Baker, R.H. and Wilkinson, G.S. (1999) “Coevolution of sperm and female reproductive tract morphology in stalk-eyed flies,” *Proceedings of the Royal Society B: Biological Sciences*, 266(1423), pp. 1041–1047. Available at: https://doi.org/10.1098/rspb.1999.0741.
Proctor, H.C. (1998) “Indirect sperm transfer in arthropods: behavioral and evolutionary trends,” Annual Review of Entomology, 43, pp. 153–174. Available at: https://doi.org/10.1146/annurev.ento.43.1.153.
Quinlan, A.R. and Hall, I.M. (2010) “BEDTools: A flexible suite of utilities for comparing genomic features,” *Bioinformatics*, 26(6), pp. 841–842. Available at: https://doi.org/10.1093/bioinformatics/btq033.
R Core Team (2017) “R: A Language and Environment for Statistical Computing.” Vienna, Austria.
Rafiqi, A.M. *et al.* (2008) “Evolutionary origin of the amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of zen,” *Proceedings of the National Academy of Sciences of the United States of America*, 105(1), pp. 234–239. Available at: https://doi.org/10.1073/pnas.0709145105.
Richardson, M.L. *et al.* (2010) “Causes and consequences of cannibalism in noncarnivorous insects,” *Annual Review of Entomology*, 55, pp. 39–53. Available at: https://doi.org/10.1146/annurev-ento-112408-085314.
Rigaud, T. and Juchault, P. (1995) “Success and failure of horizontal transfers of feminizing Wolbachia endosymbionts in woodlice,” *Journal of Evolutionary Biology*, 8(2), pp. 249–255. Available at: https://doi.org/10.1046/j.1420-9101.1995.8020249.x.
Riparbelli, M.G., Giordano, R. and Callaini, G. (2006) “Centrosome inheritance in the parthenogenetic egg of the collembolan Folsomia candida,” *Cell and Tissue Research*, 326(3), pp. 861–872. Available at: https://doi.org/10.1007/s00441-006-0253-x.
Rodriguez-R, L.M. *et al.* (2024) “An ANI gap within bacterial species that advances the definitions of intra-species units,” *mBio*, 15(1), pp. 1–14. Available at: https://doi.org/10.1128/mbio.02696-23.
Roelofs, D. *et al.* (2020) “Multi-faceted analysis provides little evidence for recurrent whole-genome duplications during hexapod evolution,” *BMC Biology*, 18(57), pp. 1–13. Available at: https://doi.org/10.1186/s12915-020-00789-1.
Ross, P.A. *et al.* (2017) “Wolbachia Infections in Aedes aegypti Differ Markedly in Their Response to Cyclical Heat Stress,” *PLoS Pathogens*, 13(1), pp. 1–17. Available at: https://doi.org/10.1371/journal.ppat.1006006.
Ross, P.A. (2024) “The use of Wolbachia for mosquito control,” in Charlwood, J.D. (ed.) Practical control of mosquito disease vectors. CABI, pp. 107–138. Available at: https://doi.org/10.1079/9781789248845.0006.
Rothberg, J.M. *et al.* (1990) “slit: An extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains,” *Genes and Development*, 4(12 A), pp. 2169–2187. Available at: https://doi.org/10.1101/gad.4.12a.2169.
Rozen, S. *et al.* (2003) “Abundant gene conversion between arms of palindromes in human and ape Y chromosomes,” *Nature*, 423(6942), pp. 873–876. Available at: https://doi.org/10.1038/nature01723.
Russell, J.E. and Stouthamer, R. (2011) “The genetics and evolution of obligate reproductive parasitism in Trichogramma pretiosum infected with parthenogenesis-inducing Wolbachia,” *Heredity*, 106(1), pp. 58–67. Available at: https://doi.org/10.1038/hdy.2010.48.
Rutherford, K. *et al.* (2000) “Artemis: Sequence visualization and annotation,” *Bioinformatics*, 16(10), pp. 944–945. Available at: https://doi.org/10.1093/bioinformatics/16.10.944.
Scarborough, C.L., Ferrari, J. and Godfray, H.C.J. (2005) “Ecology: Aphid protected from pathogen by endosymbiont,” *Science*, 310(5755), p. 1781. Available at: https://doi.org/10.1126/science.1120180.

Schaller, F. (1971) “Indirect sperm transfer by soil arthropods,” Annual Review of Entomology, 16, pp. 407–446. Available at: https://doi.org/10.1146/annurev.en.16.010171.002203.
Schmidt-Ott, U. (2000) “The amnioserosa is an apomorphic character of cyclorrhaphan flies,” *Development Genes and Evolution*, 210(7), pp. 373–376. Available at: https://doi.org/10.1007/s004270000068.
Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) “NIH Image to ImageJ: 25 years of image analysis,” *Nature Methods*, 9(7), pp. 671–675. Available at: https://doi.org/10.1038/nmeth.2089.
Sharma, A.K. and Som, A. (2023) “Assigning new supergroups V and W to the Wolbachia diversity,” *Bioinformation*, 19(3), pp. 336–340. Available at: https://doi.org/10.6026/97320630019336.
Shigenobu, S. *et al.* (2000) “Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS,” *Nature*, 407(6800), pp. 81–86. Available at: https://doi.org/10.1038/35024074.
Simakov, O. *et al.* (2022) “Deeply conserved synteny and the evolution of metazoan chromosomes,” *Science Advances*, 8(5). Available at: https://doi.org/10.1126/sciadv.abi5884.
Simão, F.A. *et al.* (2015) “BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs,” *Bioinformatics*, 31(19), pp. 3210–3212. Available at: https://doi.org/10.1093/bioinformatics/btv351.
Siozios, S. *et al.* (2013) “The Diversity and Evolution of Wolbachia Ankyrin Repeat Domain Genes,” *PLoS ONE*, 8(2). Available at: https://doi.org/10.1371/journal.pone.0055390.
Siozios, S. *et al.* (2024) “Genome dynamics across the evolutionary transition to endosymbiosis,” *Current Biology*, 34(24), pp. 5659-5670.e7. Available at: https://doi.org/10.1016/j.cub.2024.10.044.
Smit, A.F.A. and Hubley, R. (2008–2015) RepeatModeler Open-1.0 [Software]. Available at: http://www.repeatmasker.org.
Smit, A.F.A., Hubley, R. and Green, P. (2013–2015) RepeatMasker Open-4.0 [Software]. Available at: http://www.repeatmasker.org.
Snider, R. (1973) “Laboratory observations on the biology of Folsomia candida (Willem) (Collembola: Isotomidae).,” *Revue D’Écologie et Biologie du Sol*, 10(1), pp. 103–124.
Sproul, J.S. *et al.* (2023) “Analyses of 600+ insect genomes reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation challenges,” *Genome Research*, 33(10), pp. 1708–1717. Available at: https://doi.org/10.1101/gr.277387.122.
Staley, J.T. (2006) “The bacterial species dilemma and the genomic-phylogenetic species concept,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1475), pp. 1899–1909. Available at: https://doi.org/10.1098/rstb.2006.1914.
Stanke, M. and Morgenstern, B. (2005) “AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints,” *Nucleic Acids Research*, 33(SUPPL. 2), pp. 465–467. Available at: https://doi.org/10.1093/nar/gki458.
Stewart, A.C., Osborne, B. and Read, T.D. (2009) “DIYA: A bacterial annotation pipeline for any genomics lab,” *Bioinformatics*, 25(7), pp. 962–963. Available at: https://doi.org/10.1093/bioinformatics/btp097.
Stouthamer, R., Breeuwer, J.A.J. and Hurst, G.D.D. (1999) “*Wolbachia Pipientis* : Microbial Manipulator of Arthropod Reproduction,” *Annual Review of Microbiology*, 53(1), pp. 71–102. Available at: https://doi.org/10.1146/annurev.micro.53.1.71.
Stouthamer, R. and Kazmer, D.J. (1994) “Cytogenetics of microbe-associated parthenogenesis and its consequences for gene flow in Trichogramma wasps,” *Heredity*, 73(3), pp. 317–327. Available at: https://doi.org/10.1038/hdy.1994.139.
Sutton, E.R. *et al.* (2014) “Comparative genome analysis of Wolbachia strain wAu,” *BMC Genomics*, 15(1), p. 928. Available at: https://doi.org/10.1186/1471-2164-15-928.
Suyama, M., Torrents, D. and Bork, P. (2006) “PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.,” *Nucleic acids research*, 34(Web Server issue), pp. W609-12. Available at: https://doi.org/10.1093/nar/gkl315.
Tao, Y. *et al.* (2023) “Comprehensive identification of mitochondrial pseudogenes (NUMTs) in the human telomere-to-telomere reference genome,” *Genes*, 14(11), p. 2092. Available at: https://doi.org/10.3390/genes14112092.
Testerink, G.J. (1982) “Strategies in energy consumption and partitioning in Collembola,” *Ecological Entomology*, 7(3), pp. 341–351. Available at: https://doi.org/10.1111/j.1365-2311.1982.tb00675.x.

Tey, B.T. *et al.* (2009) “Production of fusion m13 phage bearing the di-sulphide constrained peptide sequence (C-WSFFSNI-C) that interacts with hepatitis B core antigen,” *African Journal of Biotechnology*, 8(2), pp. 268–273.
Timmermans, M.J.T.N. and Ellers, J. (2009) “Wolbachia endosymbiont is essential for egg hatching in a parthenogenetic arthropod,” *Evolutionary Ecology*, 23(6), pp. 931–942. Available at: https://doi.org/10.1007/s10682-008-9282-0.
Torres, D.E. *et al.* (2020) “Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model,” *Fungal Biology Reviews*, 34(3), pp. 136–143. Available at: https://doi.org/10.1016/j.fbr.2020.07.001.
Trombetta, B. and Cruciani, F. (2017) “Y chromosome palindromes and gene conversion,” *Human Genetics*, 136(5), pp. 605–619. Available at: https://doi.org/10.1007/s00439-017-1777-8.
Tulgetske, G.M. (2010) Investigations into the mechanisms of Wolbachia induced parthenogenesis and sex determination in the parasitoid wasp, Trichogramma. PhD thesis. University of California, Riverside. Available at: https://escholarship.org/uc/item/52w0b481.
Tully, T. *et al.* (2006) “Two major evolutionary lineages revealed by molecular phylogeny in the parthenogenetic collembola species Folsomia candida,” *Pedobiologia*, 50(2), pp. 95–104. Available at: https://doi.org/10.1016/j.pedobi.2005.11.003.
Tully, T. and Potapov, M. (2015) “Intraspecific Phenotypic Variation and Morphological Divergence of Strains of Folsomia candida (Willem) (Collembola: Isotomidae), the ‘Standard’ Test Springtaill,” *PLoS ONE*. Edited by P. Shaw, 10(9), p. e0136047. Available at: https://doi.org/10.1371/journal.pone.0136047.
Tvedte, E.S., Logsdon, J.M. and Forbes, A.A. (2019) “Sex loss in insects: causes of asexuality and consequences for genomes,” *Current Opinion in Insect Science*, 31, pp. 77–83. Available at: https://doi.org/10.1016/j.cois.2018.11.007.
Vanneste, K., Van de Peer, Y. and Maere, S. (2013) “Inference of Genome Duplications from Age Distributions Revisited,” *Molecular Biology and Evolution*, 30(1), pp. 177–190. Available at: https://doi.org/10.1093/molbev/mss214.
Vargas-Chávez, C. *et al.* (2025) “An episodic burst of massive genomic rearrangements and the origin of non-marine annelids,” *Nature Ecology and Evolution*, 9(7), pp. 1263–1279. Available at: https://doi.org/10.1038/s41559-025-02728-1.
Vavre, F. *et al.* (1999) “Phylogenetic evidence for horizontal transmission of Wolbachia in host- parasitoid associations,” *Molecular Biology and Evolution*, 16(12), pp. 1711–1723. Available at: https://doi.org/10.1093/oxfordjournals.molbev.a026084.
Verhulst, E.C., Pannebakker, B.A. and Geuverink, E. (2023) “Variation in sex determination mechanisms may constrain parthenogenesis-induction by endosymbionts in haplodiploid systems,” *Current Opinion in Insect Science*, 56(April), p. 101023. Available at: https://doi.org/10.1016/j.cois.2023.101023.
Visser, B. *et al.* (2010) “Loss of lipid synthesis as an evolutionary consequence of a parasitic lifestyle,” *Proceedings of the National Academy of Sciences of the United States of America*, 107(19), pp. 8677–8682. Available at: https://doi.org/10.1073/pnas.1001744107.
Visser, B. and Ellers, J. (2008) “Lack of lipogenesis in parasitoids: A review of physiological mechanisms and evolutionary implications,” *Journal of Insect Physiology*, 54(9), pp. 1315–1322. Available at: https://doi.org/10.1016/j.jinsphys.2008.07.014.
Waldorf, E. (1975) “Grooming Through the Reproductive Cycle in Male Sinella Coeca (Collembola: Entomobryidae),” *Psyche (New York)*, 82(3–4), pp. 359–365. Available at: https://doi.org/10.1155/1975/56989.
Waldorf, E.S. (1971) “The reproductive biology of Sinella curviseta (Collembola: Entomobryidae) in laboratory culture,” Revue d’écologie et de biologie du sol, 8(3), pp. 451–463.
Walter, D.E. and Proctor, H.C. (2013) Mites: ecology, evolution & behaviour: life at a microscale. 2nd edn. Dordrecht: Springer. Available at: https://doi.org/10.1007/978-94-007-7164-2.
Wang, Y. *et al.* (2012) “MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity,” *Nucleic Acids Research*, 40(7). Available at: https://doi.org/10.1093/nar/gkr1293.

Weldon, S.R., Strand, M.R. and Oliver, K.M. (2013) “Phage loss and the breakdown of a defensive symbiosis in aphids,” *Proceedings of the Royal Society B: Biological Sciences*, 280(1751), p. 20122103. Available at: https://doi.org/10.1098/rspb.2012.2103.

Werren, J.H. (1997) “BIOLOGY OF *WOLBACHIA*,” *Annual Review of Entomology*, 42(1), pp. 587–609. Available at: https://doi.org/10.1146/annurev.ento.42.1.587.

Werren, J.H., Baldo, L. and Clark, M.E. (2008) “Wolbachia: Master manipulators of invertebrate biology,” *Nature Reviews Microbiology*, 6(10), pp. 741–751. Available at: https://doi.org/10.1038/nrmicro1969.

Wolf, J.B.W. *et al.* (2009) “Nonlinear Dynamics of Nonsynonymous (dN) and Synonymous (dS) Substitution Rates Affects Inference of Selection,” *Genome Biology and Evolution*, 1, pp. 308–319. Available at: https://doi.org/10.1093/gbe/evp030.

Womack, M.C., Fiero, T.S. and Hoke, K.L. (2018) “Trait independence primes convergent trait loss,” *Evolution*, 72(3), pp. 679–687. Available at: https://doi.org/10.1111/evo.13442.

Wu, C. *et al.* (2017) “Analysis of the genome of the New Zealand giant collembolan (Holacanthella duospinosa) sheds light on hexapod evolution,” *BMC Genomics*, 18(1), p. 795. Available at: https://doi.org/10.1186/s12864-017-4197-1.

Wu, L. and Hickson, I.D. (2006) “DNA Helicases Required for Homologous Recombination and Repair of Damaged Replication Forks,” *Annual Review of Genetics*, 40(1), pp. 279–306. Available at: https://doi.org/10.1146/annurev.genet.40.110405.090636.

Wu, M. *et al.* (2004) “Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: A streamlined genome overrun by mobile genetic elements,” *PLoS Biology*, 2(3), p. e69. Available at: https://doi.org/10.1371/journal.pbio.0020069.

Xi, Z. *et al.* (2008) “Genome-wide analysis of the interaction between the endosymbiotic bacterium Wolbachia and its Drosophila host,” *BMC Genomics*, 9(1), p. 1. Available at: https://doi.org/10.1186/1471-2164-9-1.

Yang, Z. (2007) “PAML 4: Phylogenetic Analysis by Maximum Likelihood,” *Molecular Biology and Evolution*, 24(8), pp. 1586–1591. Available at: https://doi.org/10.1093/molbev/msm088.

Zabalou, S. *et al.* (2008) “Multiple rescue factors within a Wolbachia strain,” *Genetics*, 178(4), pp. 2145–2160. Available at: https://doi.org/10.1534/genetics.107.086488.

Zchori-Fein, E. *et al.* (2001) “A newly discovered bacterium associated with parthenogenesis and a change in host selection behavior in parasitoid wasps,” *Proceedings of the National Academy of Sciences*, 98(22), pp. 12555–12560. Available at: https://doi.org/10.1073/pnas.221467498.

Zchori-Fein, E., Roush, R.T. and Hunter, M.S. (1992) “Male production induced by antibiotic treatment in Encarsia formosa (Hymenoptera: Aphelinidae), an asexual species,” *Experientia*, 48(1), pp. 102–105. Available at: https://doi.org/10.1007/BF01923619.

Zhan, Z. et al. (2026) “Haplotype-resolved genome sequences of a springtail species, Folsomia candida (Collembola: Isotomidae),” Scientific Data. Available at: https://doi.org/10.1038/s41597-026-07038-0.

Zizzari, Z. V., van Straalen, N.M. and Ellers, J. (2016) “Transgenerational effects of nutrition are different for sons and daughters,” *Journal of evolutionary biology*, 29(7), pp. 1317–1327. Available at: https://doi.org/10.1111/jeb.12872.

Zizzari, Z.V. *et al.* (2009) “Female preference and fitness benefits of mate choice in a species with dissociated sperm transfer,” *Animal Behaviour*, 78(5), pp. 1261–1267. Available at: https://doi.org/10.1016/j.anbehav.2009.08.014.

Zizzari, Z.V. *et al.* (2017) “Love at first sniff: a spermatophore-associated pheromone mediates partner attraction in a collembolan species,” *Animal Behaviour*, 124, pp. 221–227. Available at: https://doi.org/10.1016/j.anbehav.2016.12.015.

Zizzari, Z.V. and Ellers, J. (2014) “Rapid shift in thermal resistance between generations through maternal heat exposure,” *Oikos*, 123(11), pp. 1365–1370. Available at: https://doi.org/10.1111/oik.01496.

Zizzari, Z.V., Jessen, A. and Koene, J.M. (2017) “Male reproductive suppression: Not a social affair,” *Current Zoology*, 63(5), pp. 573–579. Available at: https://doi.org/10.1093/cz/zow089.

Zizzari, Z.V., van Straalen, N.M. and Ellers, J. (2013) “Male-male competition leads to less abundant but more attractive sperm,” *Biology Letters*, 9(6), pp. 20130762–20130762. Available at: https://doi.org/10.1098/rsbl.2013.0762.

Zug, R. and Hammerstein, P. (2012) “Still a host of hosts for Wolbachia: Analysis of recent data suggests that 40% of terrestrial arthropod species are infected,” *PLoS ONE*, 7(6), pp. 7–9. Available at: https://doi.org/10.1371/journal.pone.0038544.

Zug, R. and Hammerstein, P. (2015) “Wolbachia and the insect immune system: What reactive oxygen species can tell us about the mechanisms of Wolbachia-host interactions,” *Frontiers in Microbiology*, 6(OCT), pp. 1–16. Available at: https://doi.org/10.3389/fmicb.2015.01201.

Zwaenepoel, A. and Van de Peer, Y. (2019) “Inference of Ancient Whole-Genome Duplications and the Evolution of Gene Duplication and Loss Rates,” *Molecular biology and evolution*, 36(7), pp. 1384–1404. Available at: https://doi.org/10.1093/molbev/msz088.

List of publications

Kampfraath, A.A., Hunting, E., Mulder, C., Breure, A., Gessner, M., Kraak, M., & Admiraal, W. (2012). DECOTAB: A multipurpose standard substrate to assess effects of litter quality on microbial decomposition and invertebrate consumption. *Freshwater Science, 31*(4), 1156–1162. https://doi.org/10.1899/12-075.1

Hunting, E., & **Kampfraath, A.A.** (2013). Contribution of bacteria to redox potential (E_h) measurements in sediments. *International Journal of Environmental Science and Technology, 10*(1), 55–62. https://doi.org/10.1007/s13762-012-0080-4

Kampfraath, A.A., Giesen, D., Van Gestel, C., & Le Lann, C. (2017). Pesticide stress on plants negatively affects parasitoid fitness through a bypass of their phytophage hosts. *Ecotoxicology, 26*(3), 383–395. https://doi.org/10.1007/s10646-017-1771-x

Faddeeva-Vakhrusheva, A., Kraaijeveld, K., Derks, M., Anvar, S., Agamennone, V., Suring, W., **Kampfraath, A.A.,** Ellers, J., Le Ngoc, G., Gestel, C., & others. (2017). Coping with living in the soil: The genome of the parthenogenetic springtail *Folsomia candida*. *BMC Genomics, 18*(1), 493. https://doi.org/10.1186/s12864-017-3852-x

Kampfraath, A.A., Klasson, L., Anvar, S., Vossen, R., Roelofs, D., Kraaijeveld, K., & Ellers, J. (2019). Genome expansion of an obligate parthenogenesis-associated *Wolbachia* poses an exception to the symbiont reduction model. *BMC Genomics, 20*(1), 106. https://doi.org/10.1186/s12864-019-5492-9

Hunting, E., Harrison, R., Bruder, A., Van Bodegom, P., Geest, H., **Kampfraath, A.A.**, Vorenhout, M., Admiraal, W., Cusell, C., & Gessner, M. (2019). Atmospheric electricity influencing biogeochemical processes in soils and sediments. *Frontiers in Physiology, 10*, 378. https://doi.org/10.3389/fphys.2019.00378

Roelofs, D., Zwaenepoel, A., Sistermans, T., Nap, J., **Kampfraath, A.A.**, Peer, Y., Ellers, J., & Kraaijeveld, K. (2020). Multi-faceted analysis provides little evidence for recurrent whole-genome duplications during hexapod evolution. *BMC Biology, 18*(1), 57. https://doi.org/10.1186/s12915-020-00789-1

Kampfraath, A.A., Dudink, T., Kraaijeveld, K., Ellers, J., & Zizzari, Z. (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

Vreman, S., Heijden, E., Ravesloot, L., Ludwig, I., Brand, J., Harders, F., **Kampfraath, A.A.**, Egberink, H., Gonzales, J., Oreshkova, N., & others. (2023). Immune responses and pathogenesis following experimental SARS-CoV-2 infection in domestic cats. *Viruses, 15*(5), 1052. https://doi.org/10.3390/v15051052

Graham, H., Most, M., **Kampfraath, A.A.**, Visser, V., Dinkla, A., Harders, F., Ruuls, R., Essen-Zandbergen, A., Esker, M., Heide, R., & others. (2024). Transmission of *Brucella canis* in a canine kennel following introduction of an infected dog. *Veterinary Microbiology, 296*, 110183. https://doi.org/10.1016/j.vetmic.2024.110183

Campos, C., Ibañez-Lligona, M., Colomer-Castell, S., Gregori, J., Garcia-Cehic, D., Andrés, C., Piñana, M., González-Sánchez, A., Rando-Segura, A., Esperalba, J., Saubí, N., Francesca Cortese, M., Tabernero, D., Rodriguez-Frias, F., Ferrer, R., Ignacio Esteban, J., Hakze-van der Honing, R., **Kampfraath, A.A.**, Vergara-Alert, J., Segalés, J., van der Poel, W., Pumarola, T., Antón & A., Quer, J. (2025). Deep sequencing reveals dual evolution of SARS-CoV-2: Insights into defective genomes from Wuhan-Hu-1 variants to Omicron subvariants. *Journal of Medical Virology, 97*(7), e70476. https://doi.org/10.1002/jmv.70476

Kampfraath, A.A., Koomen, J., Dijkstra, T., Koets, A., & Esker, M. (2025). Quick screening for *Brucella* virulence genes with the Bat-VFDB: *Brucella* addition to the Virulence Factor DataBase. *Microbiology Resource Announcements, 14*(11), e00730-25. https://doi.org/10.1128/mra.00730-25

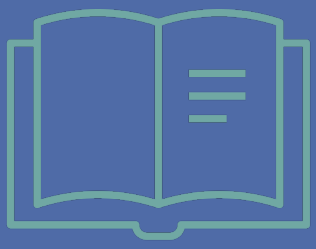
Freitas Costa, E., **Kampfraath, A.A.**, Schokker, D., Voort, M., Pijnacker, R., Vos, C., Evers, E., Bossers, A., Gonzales, J., & Pacholewicz, E. (2026). Next generation risk assessment: A proof of concept for the integration of genomic data on cold tolerance into quantitative microbial risk assessment for *Campylobacter jejuni* in poultry meat. *Microbial Risk Analysis, 31*, 100365. https://doi.org/10.1016/j.mran.2026.100365

Kannekens-Jager, M., **Kampfraath, A.A.**, Kaandorp, C., Dinkla, A., Harders, F., Ravesloot, L., Jacobs, P., Brand, J., Weerts, E., Janssen, S., & others. (2026). Impact of *Coxiella burnetii*-associated abortions in Finnish forest reindeer within a zoo on public health appeared negligible succeeding mitigation strategies. *Frontiers in Veterinary Science, 13*, 1882481. https://doi.org/10.3389/fvets.2026.1882481

Allaart, J., **Kampfraath, A.A.**, Harders, F., Voorbergen-Laarman, M., Gelderen, B., Fouz, B., & Engelsma, M. (2026). Characterization of *Edwardsiella* strains isolated from fish in the Netherlands (1995–2020) shows a relationship between *Edwardsiella* species and source. *Pathogens, 15*(9), 911. https://doi.org/10.3390/pathogens15090911

Azevedo, M., Nunes, A., Lemos, M.-L., Costa, P.M., Amaro, A., Clemente, L., Leão, C., Ramos, S., Santos, I., Joensen, K.G., Ragimbeau, C., de Baets, L., Mossong, J., Banerji, S., Flieger, A., Espadinha, D., Byrne, W., van den Berg, O., **Kampfraath, A.A.**, Jehanne, Q., Lehours, P., Gomes, J.P., & Oleastro, M. (2026). Multisectoral emergence of multidrug-resistant *Campylobacter coli* sequence type 10042 lineage, Europe, 2018–2025. *Emerging Infectious Diseases, 32*(10), 1645-1656. https://doi.org/10.3201/eid3210.260512

SUMMARY







DKJR

Summary

The evolutionary transition from sexual to asexual reproduction represents one of the most profound lifestyle shifts across the tree of life, causing large shifts in selective pressures on phenotypic and genotypic traits. These shifts in pressures could lead to the decay and disappearance of sexual traits no longer needed for asexual reproduction. When asexual reproduction is induced by a symbiotic organism, that organism may compensate for host functions that become redundant after the loss of sexual reproduction. This process, in which the symbiont replaces or maintains functions formerly supported by sexual traits or processes, is termed compensated trait loss. This thesis investigates the multifaceted consequences of the transition to asexuality, with particular emphasis on compensated trait loss, using the springtail *Folsomia candida* and its endosymbiont *Wolbachia* as a study system. *Wolbachia* is known for its reproductive manipulations and there are strong indications it is involved in the asexual reproduction of *F. candida*. Its transition to reproduction through parthenogenesis was examined through phenotypic, symbiotic, and genomic lenses, to get a full scope of the involved processes and genomic consequences.

The investigation began at the phenotypic level, focusing on the fate of male-specific traits that become redundant in parthenogenetic lineages. Through a comparative analysis of sexual and asexual *F. candida* populations, it was demonstrated that several key reproductive traits, specifically sperm DNA quantity and the olfactory attractiveness of spermatophores, underwent significant degradation in asexual lines. The observed patterns of decay, which varied between lineages, are highly consistent with the neutral mutation accumulation theory, suggesting that once the selective pressure to maintain these traits is removed, stochastic loss-of-function mutations become the primary driver of phenotypic deterioration. It can be argued that this is a form of compensated trait loss, given the assumption that *Wolbachia* is allowing parthenogenetic reproduction, which would compensate for these phenotypic traits that are now lost.

Building upon this understanding of trait decay, the research then turned to the endosymbiont driving this reproductive shift: the *Wolbachia* strain *w*Fol. While the transition to asexuality is often linked to the stability of an endosymbiotic infection, our experiments with transinfection and heat stress revealed a complex interplay between host immunity and symbiont persistence. Notably, we observed that while high temperatures directly reduced *Wolbachia* titres, they did not necessarily trigger the lytic cycle of the associated bacteriophage WO. Furthermore, the difficulty in establishing stable infections in sexual lineages suggests that the host's innate immune response may act as a significant barrier to horizontal symbiont transfer, potentially reinforcing the evolutionary boundaries between sexual and asexual populations.

Subsequently, the exploration proceeded at the genomic level where it challenged established paradigms regarding endosymbiont evolution. Contrary to the long-standing symbiont reduction model, which predicts genome streamlining in obligate relationships, we

discovered that the *w*FoI genome is exceptionally large and characterized by massive expansions of phage-derived regions. This unexpected expansion, accompanied by an enrichment of DNA-repair and ankyrin-repeat genes, suggests a mode of genome preservation rather than reduction. This pattern of genomic stability in the endosymbiont was mirrored in the host; our comparative analysis of *F. candida* lineages revealed that while sexual populations maintain a streamlined architecture, asexual lineages exhibit significant genomic expansion. This is characterized by an increased accumulation and retention of repetitive elements, segmental duplications, and mitochondrial DNA remnants.

Ultimately, this thesis hypothesises that the transition to parthenogenesis in *F. candida* is not merely a process of trait loss, but one of complex genomic restructuring. The decay of sexual traits is compensated for by a unique evolutionary strategy in asexual lineages: the adoption of a genome architecture optimized for the accumulation and preservation of genetic material. This shift from purging to conservation likely serves as an adaptive mechanism to navigate the lack of sexual recombination, providing a window into the broader dynamics of genome evolution during major life-history transitions.

Samenvatting

De overgang van seksuele naar aseksuele voortplanting is een van de ingrijpendste veranderingen in de evolutie van organismen. Zo'n overgang verandert de selectiedruk op eigenschappen die betrokken zijn bij voortplanting. Seksuele kenmerken die niet langer nodig zijn, kunnen daardoor geleidelijk achteruitgaan en uiteindelijk verdwijnen. Wanneer aseksuele voortplanting wordt veroorzaakt door een endosymbiotisch organisme (een organisme dat in of binnen de cellen van een ander organisme leeft), kan dit organisme functies van de gastheer overnemen of compenseren die door de overgang naar aseksualiteit overbodig zijn geworden. Dit verschijnsel staat bekend als *gecompenseerd eigenschapsverlies*.

Dit proefschrift onderzoekt de gevolgen van de overgang naar aseksuele voortplanting, met bijzondere aandacht voor gecompenseerd eigenschapsverlies. Als modelsysteem is de springstaart *Folsomia candida* gebruikt, samen met diens endosymbiont *Wolbachia*. *Wolbachia* is een bacterie die bekendstaat om het beïnvloeden van de voortplanting van zijn gastheren. Er zijn sterke aanwijzingen dat deze bacterie ook de aseksuele voortplanting van *F. candida* veroorzaakt. In dit proefschrift wordt de overgang naar parthenogenese (een specifieke vorm van aseksuele voortplanting vanuit onbevruchte eicellen) onderzocht vanuit drie invalshoeken: het fenotype, de symbiose en het genoom. Samen geven deze perspectieven inzicht in de processen achter deze reproductieve omslag en in de gevolgen ervan voor de genomen van zowel gastheer als symbiont.

Het onderzoek begon met de vraag wat er gebeurt met mannelijke kenmerken die in parthenogenetische lijnen niet langer functioneel of noodzakelijk zijn. Een vergelijking van seksuele en aseksuele populaties van *F. candida* liet zien dat verschillende belangrijke reproductieve eigenschappen in aseksuele lijnen sterk zijn achteruitgegaan. Dit gold met name voor de hoeveelheid sperma in de spermatofoor (bolletje met sperma dat kan worden opgepakt door het vrouwtje om haar eitjes te bevruchten) en voor de aantrekkingskracht van spermatoforen op basis van geurstoffen. De mate en het patroon van deze achteruitgang verschilden tussen lijnen. Dit past bij de theorie van neutrale mutatieaccumulatie, waarbij zodra de natuurlijke selectie op een eigenschap niet langer in stand wordt gehouden willekeurige mutaties kunnen ophopen en uiteindelijk functieverlies optreedt. Wanneer *Wolbachia* inderdaad de parthenogenetische voortplanting mogelijk maakt, kan het verdwijnen van deze mannelijke eigenschappen worden gezien als gecompenseerd eigenschapsverlies.

Vervolgens richtte het onderzoek zich op de endosymbiont die waarschijnlijk verantwoordelijk is voor deze reproductieve omslag: de *Wolbachia*-stam wFol. De overgang naar aseksualiteit wordt vaak gekoppeld aan een stabiele, langdurige endosymbiotische infectie. Experimenten met transinfectie en

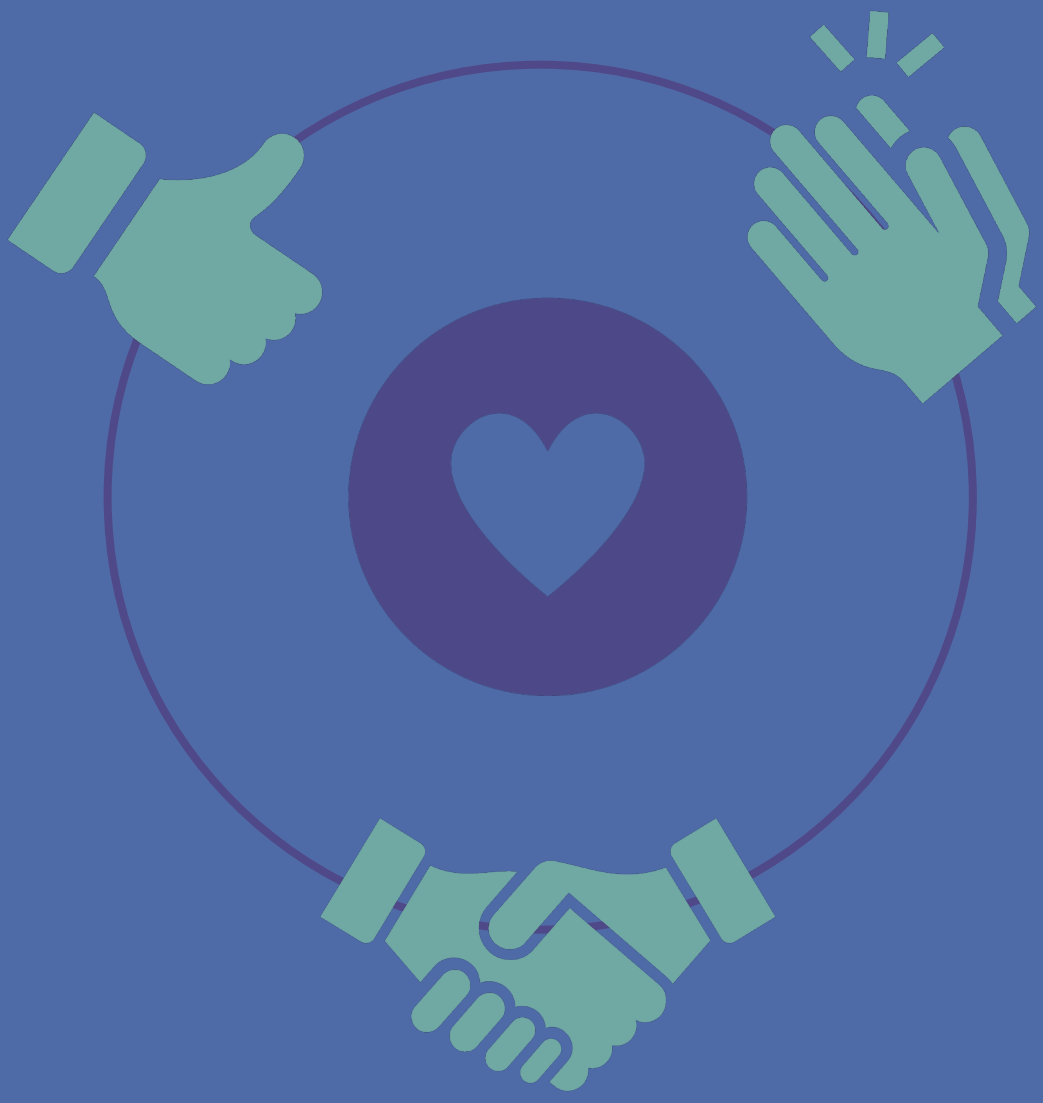
hittestress lieten echter zien dat de relatie tussen gastheer en symbiont complexer is. Hoge temperaturen verlaagden weliswaar het aantal *Wolbachia*-bacteriën, maar activeerden niet de lytische cyclus van de bijbehorende bacteriofaag WO. Daarnaast bleek het moeilijk om stabiele *Wolbachia*-infecties tot stand te brengen in seksuele lijnen van *F. candida*. Dit wijst erop dat de aangeboren afweer van de gastheer een belangrijke belemmering kan vormen voor horizontale overdracht van de symbiont. Daardoor kan dit mechanisme ook bijdragen aan het in stand houden van de evolutionaire scheiding tussen seksuele en aseksuele populaties.

Op genomisch niveau laat dit proefschrift zien dat de evolutie van endosymbionten niet altijd verloopt volgens het klassieke patroon van genoomreductie. Bij obligate symbiose wordt vaak verwacht dat een symbiontgenoom kleiner en eenvoudiger wordt doordat overbodige genen verloren gaan. Het genoom van wFol bleek echter uitzonderlijk groot. Dit genoom bevat sterke uitbreidingen van regio's die zijn afgeleid van bacteriofagen. Ook is het verrijkt met genen die betrokken zijn bij DNA-herstel. Deze kenmerken wijzen eerder op het behoud van genetisch materiaal dan op vergaande genoomreductie.

Een vergelijkbaar patroon is zichtbaar in de gastheer. Seksuele populaties van *F. candida* behouden een relatief gestroomlijnde genoomarchitectuur, terwijl aseksuele lijnen juist een aanzienlijke uitbreiding van hun genoom vertonen. Die uitbreiding gaat samen met een grotere ophoping en retentie van repetitieve DNA-elementen, duplicatie van genoomsegmenten en resten van mitochondriaal DNA. Zowel in de symbiont als in de aseksuele gastheer lijkt genoombehoud en -uitbreiding dus een belangrijker rol te spelen dan genoomstroomlijning.

Samengevat laat dit proefschrift zien dat de overgang naar parthenogenese in *F. candida* niet alleen leidt tot het verlies van seksuele eigenschappen. De overgang gaat ook gepaard met ingrijpende veranderingen in de relatie tussen gastheer en symbiont en met een grootschalige herstructurering van beide genomen. In aseksuele lijnen lijkt het verlies van seksuele functies te worden opgevangen door *Wolbachia* en door een genoomarchitectuur die de accumulatie en het behoud van genetisch materiaal bevordert. Deze verschuiving van stroomlijnen naar behoud en uitbreiding kan mogelijk helpen om de gevolgen van het ontbreken van seksuele recombinatie op te vangen. Daarmee levert dit onderzoek nieuwe inzichten op in de evolutie van genomen tijdens grote veranderingen in voortplantingswijze en levenscyclus.

ACKNOWLEDGEMENTS



DKJR

Acknowledgements

They say it takes a village to raise a child, while that is probably true, it also takes a village to complete a thesis. Especially when you take your sweet time, like I did, there are many people to thank. Where to start?

Of course, before you start doing research you need to acquire the skills to do so. For this I would like to thank all the professors and lecturers at the UvA and more specifically the supervisors of my internships, which felt as the real first steps into conducting science. So, Marino and Michiel thanks for supervising my bachelor internship and a big thanks to Ellard, Harm, Michiel, Daniel, Cecile and Kees for supervising my master internships. It is during these projects that I found my passion for research and learned how nice it could be to completely lose yourself in finding out a small detail by meticulously setting up experiments. Thanks for letting me stand on your shoulders!

Then to my promotors, at first, dear Jacintha, thanks a lot for putting your faith in me when offering me the position for this doctorate. It has been a great adventure, although at times, over bumpy roads. Therefore, an even greater thanks for all the patience and believing in a successful end of this trajectory even at times I did not believe in it myself anymore. Without your ongoing support I would have never been able to complete this thesis.
And dear Ken, thanks a lot for all the guidance during my research. When the focus became more on bioinformatics you introduced me to this field. That I am very grateful for, as it now became a core part of my job, something I never thought would happen when you first showed me some bash commands in a Linux terminal ;). Also, our trip to Namen was one of the highlights of my PhD.
In addition, I would like to thank the committee Eric, Matty, Yuval, Bertanne and Karine for taking the time to evaluate my thesis carefully.

A PhD student in the lab is only as good as the technicians supporting him, luckily, we had a great team of technicians that helped supporting me: Rudo, Riet and Janine. My thanks go especially to Janine. You helped a lot in setting up all the molecular work, from classical DNA extractions to Q-PCR you always knew how to tackle the challenge! Moreover, the way you document your work is admirable, even after 10 years you could find the plasmid used in our experiments without a problem! The department will have a hard time replacing you after you retire.

There were also other mentors that took me under their wings and shared their craft with me. To Gerard thanks a lot for teaching me about teaching and fieldwork on Schier. It was a joy to see such experience at work, and I will always look back in

fondness to that fieldwork course-week. To Seth and Sarah, my gratitude for sharing your knowledge on *Wolbachia* phages with me. To Lisa, thanks a lot for inviting me to Sweden for three weeks and teaching me how to annotate genomes properly. The insights I gained during that time were invaluable for completing my thesis and still are in my current job.

I would also like to thank the co-authors of the chapters in this thesis I did not mention already. Dick, Valentina, Yahya, Rolf, Eef and Sandra many thanks for all the attention and work you put into these manuscripts, it was a pleasure to work with all of you, hopefully there will be more collaborations in the future!

Supervising students was part of the journey as well. I had the privilege to supervise two excellent students during my PhD. Dear Jitske and Tjeerd, thanks a lot for your contributions to my research. It was my pleasure to take my turn in sharing my knowledge and guide you in your first steps of becoming fine biologists.

Of course, also my roommates at the VU deserve a word of thanks. Tjalf, Anna, Mima, Valeria, Anne, Matías and Giang thank you all for making our office a nice place to go to and sharing good scientific discussions, a laugh or a cup of tea.

Also, my current colleagues at WBVR of the bioinfo group and beyond: Dirkjan, Timo, Amirreza, Frank, Albert, Ewa, Eduardo, Jose, Heather, Manouk, Marleen, Miriam, Marielle, Jeroen, Marc, Matthijn, Valerio and others, deserve a big thanks for supporting me in finishing my thesis. And especially Alex who gently nudged me into completing this thesis and to Annemarie for motivating me to work in bigger blocks on my thesis, if not, it would not have been completed by now.

Of course, a special thanks goes out to my paranymphs as well. Dear Oscar, it has been a great pleasure to have been on this journey together. Both for nice work-related trips in the field and fruitful discussions as well as the many extracurricular activities that usually involved birds, beer or whiskey.
Dear Wendy, thanks for all the advice over the years, the good discussions on statistics and biological theories and the walks when we would both be stuck with our thesis. I'm very grateful to have the two of you as my paranymphs and will always cherish all our long nights at "de Stelling".

Talking about "de Stelling", a big shout out to Gyrinus for running this pub on Wednesday and Fridays, it has been the heart of the social network at the VU. Also, thanks to all the nice colleagues that came along for drinks after work Maartje, Jeroen, Wouter, Mirte, Wouter, Mark, Matt, Lee, Yumi, Valeria, Francesco, Estefania, Anouk, Claudia, Ruben, Simon, Estafania, Ralph, Nico, Joris, Toby, Judith, Andrew and many more. Thank you all for the great moments and the great atmosphere at the VU!

Apart from on the VU, I also got a lot of support or necessary distractions from my other friends. Thanks, Kiboe-boys: Koen, Nandor and Jaro for sticking with me for so long and doing great projects together, as well as the best gaming nights! Thanks, Klaverjassers: Alex, Wouter, Nico and Seth for all the card nights with humorous remarks to put things in perspective! Thanks Biodegradables: Alaleh, Yakhma, Mark, Thomas, Anne, Rogier, Ana and Julia for all the nice times together, may we have many more! Thanks to the partners that joined our lunches and became good friends Roy, Vincent and Henrieke, it's a joy you joined the gang! And a very special thank you to Tijmen and Marchien, who have always been there to listen. You have heard more than your fair share of complaints about my thesis over the years; I am glad that, from now on, I can bother you with other problems instead ;).

Un ringraziamento speciale a: Antonella, Paolo, Claudia, Federico e Sergio per la calorosa accoglienza nella loro famiglia, per aver condiviso la loro cultura e la loro cucina e per le meravigliose vacanze in Italia.

I would also like to thank my family: Pake, Omi, Katerina, Robin, Faye, Idske, Uli, Moritz, Kiki, Paul, Anne-Gerda, Paul, Laurine, Carolien, Frank, Emma, Floor, Chris, Franka, Jan-Chris and Eliza for always cheering me on, showing interest to my thesis and also for sometimes not asking about the status of my thesis. This meant more to me than I can express. Of course, also thanks for all the great North Sea Jazz weekends to the once who joined and a special thanks goes out to my uncle Siebe, who also taught me a lot of DIY skills that came in handy more than once during the set-up of experiments.

My parents, Karin and Dré, I would like to thank for the warm household they created for me and my sisters and for giving me the opportunity to pursue my scientific career. Also, my sisters Leanne and Marije and their families, Jackie, James, Jurriaan and Stella I would like to thank for all the support over the years. The loss of our father hit us all hard, fortunately, he's still here in all that he has taught us.

And last, but not least, dear Monica, thank you for being with me throughout this whole journey, without you I would not have never made it. From the moment we met during our master till now, it has been a privilege to have you be my side. Our shared love for science, art and colour resonates throughout this whole thesis. I also want to thank you for the wonderful home we build together and most of all for the greatest gift you could even give me: Alessandro. And Alessandro, I thank for opening new ways of looking onto the world and taking me along your exploration and amazement of everyday life. Seeing the sparks of excitement of learning new things, gave me the strength and inspiration to fulfil this thesis. I love you both very deeply, thanks for everything!

AUTHOR



DKJR

On the author

Andries Augustus Kampfraath, to most known as Dré, was born on 11 March 1989 in Amsterdam, the Netherlands. From an early age, he developed a strong interest in animals and the natural world, which later led him to study Biology at the University of Amsterdam. He obtained his BSc in Biology in 2010, followed by an MSc in Biological Sciences, with a specialisation in Limnology and Oceanography, in 2013.

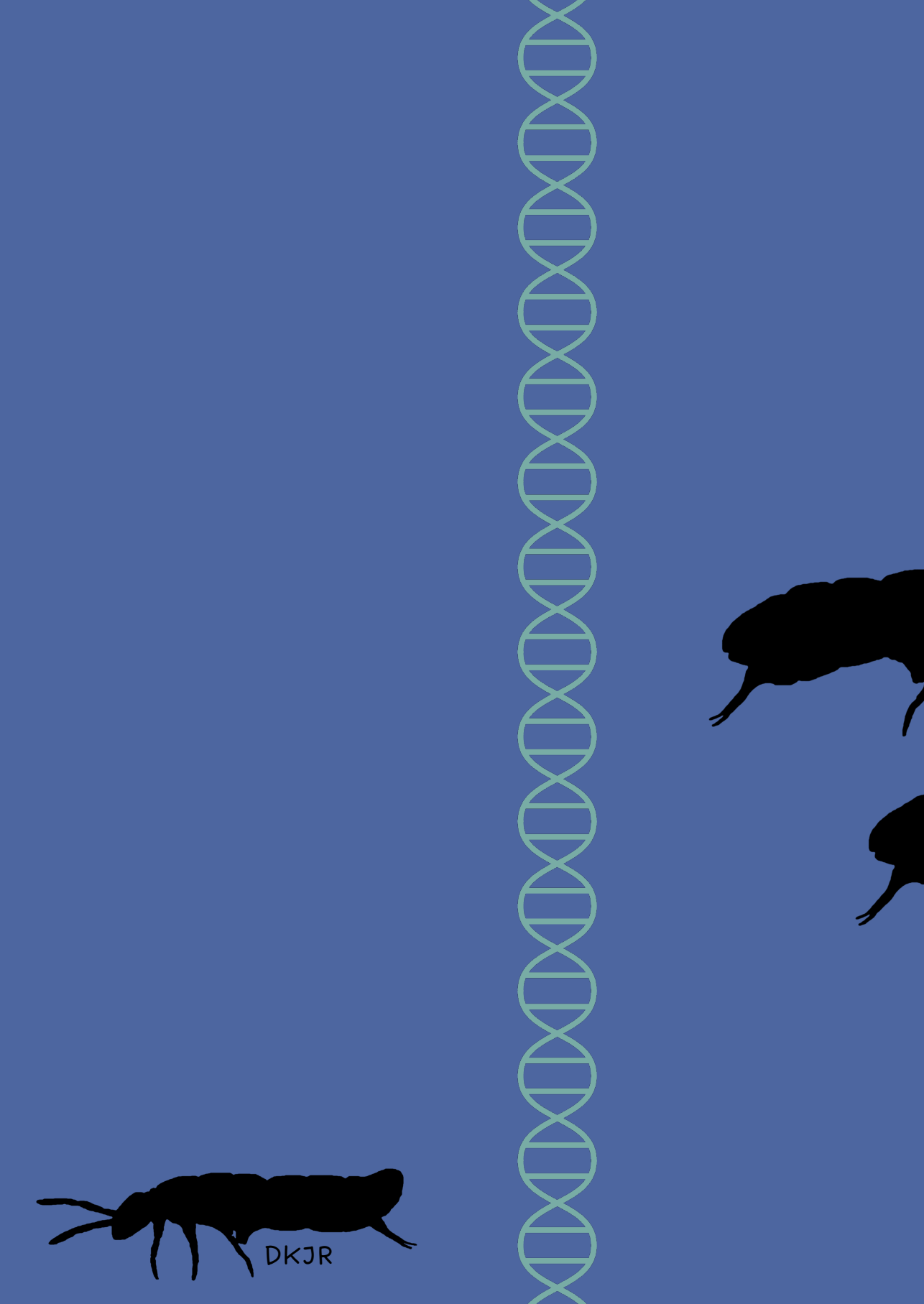
During these studies, Dré developed a broad interest in ecology, environmental toxicology, species interactions, and evolutionary biology. Within his master's internships, he investigated how invertebrate functional traits and solar irradiation influence decomposition and microbial communities in sediments, and how lipophilic organic pollutants affect tritrophic interactions among plants, herbivores, and parasitoid wasps. These projects fostered his interest in complex ecological systems and in the evolutionary and environmental processes that shape interactions among organisms.

In 2013, Dré started his PhD research at the Department of Animal Ecology of Vrije Universiteit Amsterdam. His doctoral work focused on the genetic and phenotypic consequences of the loss of sexual reproduction in the soil-dwelling springtail *Folsomia candida*, with particular attention to the role of its endosymbiont, *Wolbachia*, resulting in this thesis. This work marked his transition into bioinformatics.

Since 2019, Dré has worked primarily as a researcher and project leader in bioinformatics at Wageningen Bioveterinary Research, alongside completing his doctoral research. There, he applies genomic and evolutionary approaches to questions with direct societal relevance, including veterinary infectious diseases, zoonoses, and outbreak dynamics.

All these combined research experiences have resulted in publications spanning ecotoxicology, soil and aquatic ecology, evolutionary genomics, host–symbiont interactions, and veterinary microbiology.

Outside the office, he is a dedicated father, an enthusiast of the North Sea Jazz Festival, and an active contributor to his local community. He volunteered for many years at Petting Zoo De Pijp, including as a member of its board, and currently serves on the board of his apartment building's homeowners' association.


DKJR