DISS. ETH NO. 27282

The price of protection: Investigating costs of defensive symbiosis in aphids

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

HEIDI KAECH

MSc ETH in Biology, ETH Zurich, Switzerland

born on 14.02.1988

citizen of Buttisholz LU, Switzerland

accepted on the recommendation of

Prof. Dr. Christoph Vorburger

Prof. Dr. Jukka Jokela

Prof. Dr. Kerry M. Oliver

2020

Table of Content

Table of Content 1

Summary 2

Zusammenfassung 4

General Introduction 6

Chapter 1 15

Triple RNA-Seq characterizes aphid gene expression in response to infection with uenqually virulent strains of the endosymbiont *Hamiltonella defensa*

Chapter 2 63

Horizontal transmission of *Hamiltonella defensa* depends on titre and haplotype

Chapter 3 91

Similar cost of *Hamiltonella defensa* in experimental and natural aphid-endosymbiont associations

General Discussion 117

Acknowledgements 124

Appendix chapter 1 125

Appendix chapter 2 142

Summary

The relationship between insects and bacteria is complex. On one hand, insects need to defend themselves against infection by pathogenic bacteria. At the same time, many insects must maintain an infection with very specific symbiotic bacteria to survive. Aphids, for example, can only live off plant sap due to their primary endosymbiont *Buchnera aphidicola.* It produces the essential amino acids that the aphids cannot extract from their nutrient-poor diet.

To understand how bacteria turn from foe to ally during the evolution of symbioses, it is of particular interest to investigate secondary endosymbioses. In such an association, a bacterium is still making its way from parasite to mutualist, and the host is not yet fully dependent on the bacterium's presence. In fact, carrying the secondary endosymbiont is costly to the host. To maintain itself in the host population, the endosymbiont must compensate its cost; either by manipulating the host's reproduction or by providing the host with fitness benefits. Endosymbionts can, for example, increase their host's resistance to enemies. Such a defensive endosymbiosis can be observed in the black bean aphid (*Aphis fabae*). If this aphid species is infected with the bacterium *Hamiltonella defensa,* it experiences protection against parasitoid wasps.

My thesis focused on the price that black bean aphids pay for this protection. In absence of parasitoids, *H. defensa* can be very costly, as it curtails the aphids' lifespan and offspring production. I explored the mechanistic basis behind this *H. defensa*-induced cost. I investigated whether 'expensive' isolates of *H. defensa* compensate part of their cost by jumping from one host to another like a pathogen. And I asked whether we can predict the costs of *H. defensa* from studying artificially created aphid-endosymbiont combinations instead of studying combinations that were found in the field.

Using a triple-RNA-Seq approach, I analysed the influence of *H. defensa'*s presence on the gene expression of the black bean aphid and its primary endosymbiont *B. aphidicola*. I looked for patterns of covariation between gene expression of the aphid and *H. defensa,* and between gene expression of *B. aphidicola* and *H. defensa.* I showed that four different isolates of *H. defensa* affect their host in distinct ways, and that the costs are imposed through interaction with the aphid rather than with *B. aphidicola.*

Next, I focused the influence of transmission mode on the endosymbiosis between aphid and *H. defensa.* Generally, symbionts can use two different transmission modes: vertical from

mother to offspring and horizontal between hosts. Horizontally transmitted symbionts are expected to be more costly, if stronger exploitation of their current host increases their titre and at the same time their chances of transmitting horizontally to the next host. On a set of 11 well-characterised *H. defensa* isolates, I showed that those isolates reaching a high titre in their host are indeed more efficient at horizontal transmission. Thus, costly isolates with a high titre may compensate part of their cost through horizontal transmission. Yet, titre alone was not enough to predict horizontal transmission success – endosymbionts might have to evolve specific adaptations to transmit effectively between hosts.

Finally, I tested whether experimental infections of naturally uninfected aphid hosts with *H. defensa* can provide a representative picture of the symbiont-induced costs seen in naturally infected aphid hosts. I compared the cost of 11 *H. defensa*-isolates in their naturally associated aphid clone to their cost in two artificially infected aphid clones. While the costs in experimental associations were within the range of natural associations, I also found that symbiont-induced costs fluctuate over time in laboratory culture. Therefore, costs estimated in the laboratory may not always be representative of the costs in the field.

Together, the three chapters shed light on different aspects of the costs of a secondary endosymbiosis: their cause, a rarely considered way to compensate for them, and the accuracy of their assessment in the laboratory.

Zusammenfassung

Die Beziehung zwischen Insekten und Bakterien ist komplex. Einerseits müssen sich Insekten gegen das Eindringen von Bakterien verteidigen, damit sie nicht einem bakteriellen Krankheitserreger zu Opfer fallen. Andererseits brauchen viele Insekten Bakterien, weil sie ohne bakterielle Symbionten nicht überleben könnten. So etwa Blattläuse: Würde ihr primärer Endosymbiont, *Buchnera aphidicola*, nicht essenzielle Aminosäuren für die Blattläuse produzieren, wären diese nicht dazu imstande, vom nährstoffarmen Pflanzensaft zu leben.

Um zu verstehen, wie während der Evolution einer Symbiose ein Bakterium vom Feind zum Freund wird, ist die Untersuchung sekundärer Endosymbiosen von besonderem Interesse. Denn in sekundären Endosymbiosen befindet sich das Bakterium noch auf dem Weg vom Parasitismus zum Mutualismus: Einerseits ist der Wirt noch nicht vollständig von der Anwesenheit des sekundären Endosymbionten abhängig, andererseits ist die Anwesenheit des Endosymbionten für den Wirt mit Kosten verbunden. Diese Kosten muss ein Endosymbiont kompensieren, damit er nicht aus der Wirt-Population verschwindet. Entweder er manipuliert die Fortpflanzung seines Wirts oder er bietet ihm Fitnessvorteile. Beispielsweise können Endosymbionten die Widerstandskraft ihres Wirts gegen natürliche Feinde erhöhen. Beobachten kann man eine solche defensive Symbiose in der Schwarzen Bohnenblattlaus (*Aphis fabae*). Ist diese Blattlausart mit dem Bakterium *Hamiltonella defensa* infiziert, so schützt sie das davor, von Schlupfwespen parasitiert zu werden.

Meine Doktorarbeit beschäftigte sich mit dem Preis, den die Schwarze Bohnenblattlaus für diesen Schutz bezahlt. In Abwesenheit von Schlupfwespen wird die Infektion mit *H. defensa* nämlich zum Nachteil: *H. defensa* verkürzt die Lebensdauer der Blattlaus und verringert die Anzahl Nachkommen, die die Blattlaus produzieren kann. Ich suchte nach einer mechanistischen Grundlage für diese Kosten, die *H. defensa* dem Wirt aufbürdet. Ich untersuchte, ob ‚teure' Isolate von *H. defensa* einen Teil ihrer Kosten kompensieren, indem sie wie ein Krankheitserreger von einem Wirt zum anderen springen. Und ich fragte, ob die Kosten von *H. defensa* in natürlichen Wirt-Endosymbiont-Kombinationen durch die Analyse künstlich erzeugter Kombinationen vorhergesagt werden können.

Ich verwendete einen Triple-RNA-Seq-Ansatz, um den Einfluss von *H. defensa* auf die Genexpression der schwarzen Bohnenblattlaus und ihres primären Endosymbionten *B. aphidicola* zu untersuchen. Dazu suchte ich nach Kovariationsmustern zwischen der

Genexpression der Blattlaus (oder des primären Endosymbionten) und der Genexpression von *H. defensa*. Ich zeigte, dass sich die untersuchten vier verschiedenen *H. defensa*-Isolate unterschiedlich auf die Genexpression des Wirts auswirkten und dass die Kosten eher durch die Wechselwirkung von *H. defensa* mit der Blattlaus als mit *B. aphidicola* entstehen.

Als Nächstes testete ich, ob sich der Übertragungsmodus von *H. defensa* auf die Kosten auswirkt, die *H. defensa* im Wirt verursacht. Im Allgemeinen können Symbionten zwei verschiedene Übertragungsmodi verwenden: vertikal von der Mutter zum Nachwuchs und horizontal zwischen zwei Wirten. Horizontal übertragene Symbionten verursachen ihrem Wirt gemäss Theorie mehr Kosten als vertikal übertragene Symbionten. Könnte es also sein, dass ‚teure' *H. defensa* durch die Ausbeutung ihres aktuellen Wirts einen höheren Titer erreichen und so häufiger horizontal übertragen werden? Anhand eines Sets von 11 gut charakterisierten *H. defensa*-Isolaten zeigte ich, dass jene Isolate, die in ihrem Wirt einen hohen Titer erreichen, sich tatsächlich effizienter horizontal übertragen. Kostspielige Isolate mit einem hohen Titer können daher einen Teil ihrer Kosten durch horizontale Übertragung kompensieren. Der Titer allein reichte jedoch nicht aus, um den Erfolg der horizontalen Übertragung vorherzusagen. Endosymbionten müssen möglicherweise spezifische Anpassungen entwickeln, um effektiv von einem Wirt zum anderen springen zu können.

Schliesslich testete ich, ob experimentelle Infektionen mit *H. defensa* ein repräsentatives Bild der Kosten liefern können, die bei natürlichen Infektionen auftreten. Dazu nutze ich 11 *H. defensa*-Isolate, die entweder mit ihrem natürlichen Wirt-Genotyp assoziiert waren oder mit einem ‚experimentellen' Wirt, zwei Blattlausklonen, die im Labor mit den *H. defensa*-Isolaten infiziert worden waren. Die Kosten, die in experimentellen Assoziationen gemessen wurden, lagen im Bereich natürlicher Assoziationen. Jedoch stellte ich auch fest, dass die Kosten von *H. defensa* über fast ein Jahrzehnt der Haltung im Labor stark schwankten. Daher sind die im Labor geschätzten Kosten möglicherweise nicht immer repräsentativ für die Kosten im Feld.

Die drei Kapitel dieser Doktorarbeit beschäftigen sich mit verschiedenen Aspekten der Kosten einer sekundären Endosymbiose. Sie suchen nach der Ursache der Kosten, erkunden eine selten in Betracht gezogene Möglichkeit zur Kompensation der Kosten, und fragen nach, ob Kosten im Labor akkurat gemessen werden können.

General Introduction

Bacterial symbionts in animals

Animals live in a bacterial world (McFall-Ngai et al. 2013): Virtually all animals are associated with bacteria, many of which cannot be cultured. Interactions between bacteria and animals span the entire range from parasitism to mutualism, up to the point that neither the animal host nor the bacterial symbiont is viable without its partner. Secondary endosymbionts of insects are somewhere in the middle – in the 'tension zone', so to speak – of the mutualism-parasitism continuum. They rely on and exploit their host, but they can also provide benefits under certain environmental conditions. As a result, they often reach an intermediate frequency in their host population based on eco-evolutionary interactions with parasitoids or other environmental factors (Ives et al. 2020). Studying the interaction and co-evolution of secondary endosymbionts with their insect hosts – as I did in this thesis with *Hamiltonella defensa* and the black bean aphid – can help us understand the evolution of microbial symbioses in animals.

Endosymbiosis in aphids

Aphids are among the most important insect pests worldwide (Dedryver et al. 2010), causing substantial crop losses by vectoring plant viruses and by feeding on plant sap. They are remarkably efficient: Clonal reproduction of parthenogenetic females during spring and summer leads to rapid population growth, while winged morphs allow dispersal over large geographic scale. Parthenogenesis can be interrupted with sexual reproduction, enabling aphids to overwinter as frost-resistant eggs that give rise to new genotypes in the following spring (Van Emden and Harrington 2007).

Like many insects, aphids rely on microbial endosymbionts for their survival. Most crucial is their association with *Buchnera aphidicola,* a γ-proteobacterium which they harbour intracellularly in specialised bacteriocytes (Douglas 1998). In finely tuned coordination, bacteriocytes and *B. aphidicola* synthesize all essential amino acids lacking from the aphid's food; sugar-rich but nutrient-poor plant sap (Baumann 1995; Brinza et al. 2009; Hansen and Moran 2011; Smith and Moran 2020). Over the at least 160 Million years of their association (Moran and Baumann 1994), the symbiosis of aphids and *B. aphidicola* has progressed to such extent that neither partner can survive without the other. While antibiotic treatment can cure aphids from *B. aphidicola,* maintenance of aposymbiotic aphids on plants is impossible,

as the aphids do not grow to normal size and hardly ever reproduce (Douglas 1998). Likewise, no one has succeeded in establishing a stable *in vitro* culture of *B. aphidicola,* whose highly reduced genome contains a meagre ~540 protein-coding genes (Shigenobu et al. 2000). The relationship between aphid and *B. aphidicola* is called a primary endosymbiosis based on four traits of *B. aphidicola*: Its universal presence in all hosts, its reduced genome, its intracellular lifestyle, and its faithful vertical transmission from aphid mothers to offspring (Douglas 2016; Wilkinson et al. 2003).

In addition to their primary endosymbiont *B. aphidicola,* aphids can associate with different secondary endosymbionts. While secondary endosymbionts also transmit vertically with high fidelity (Dykstra et al. 2014) and can be found intracellularly in sheath cells adjacent to the bacteriocytes, they also occur in the host's hemolymph (Fukatsu et al. 2000; Moran et al. 2005b; Sandström et al. 2001) and do not go to fixation in the host population (Smith et al. 2015; Vorburger and Rouchet 2016). Even though the nine known secondary endosymbionts of aphids (*Arsenophonus*, *Hamiltonella defensa, Regiella insecticola, Rickettsia, Rickettsiella, Serratia symbiotica, Spiroplasma, PAXS* and *Wolbachia)* usually cannot live independently from their host, the aphids do not require them for their survival. In fact, the secondary endosymbionts' spread has been shown to be constrained by costs that they inflict on the host's lifespan, development or reproduction (Cayetano et al. 2015; Jamin and Vorburger 2019; Leybourne et al. 2020), on competitiveness in population cages (Dykstra et al. 2014; Oliver et al. 2008; Rossbacher and Vorburger 2020), or on behaviour (Dion et al. 2011; Polin et al. 2014).

Vertically transmitted endosymbionts generally use two strategies to persist or spread in a host population: through reproductive manipulation or by providing fitness benefits. An example for an endosymbiont that manipulates the reproduction of its hosts in a way that favours its own spread is *Wolbachia,* probably the most widespread secondary endosymbiont of insects, or *Spiroplasma*, which has been estimated to infect an estimated 4-7% of arthropod species (Duron et al. 2008; Hilgenboecker et al. 2008; Werren et al. 2008). Increasing the host's fitness can entail providing a defensive function – some *Spiroplasma* protect pea aphids against entomopathogenic fungi or parasitoid wasps (Frago et al. 2017; Łukasik et al. 2013; McLean et al. 2020) – or conveying other benefits such as increased resistance to heat stress (Montllor et al. 2002).

The main secondary endosymbiont studied in this thesis is *Hamiltonella defensa* (Moran et al. 2005b). Just like its name implies, *H. defensa* protects several aphid species – for example the pea aphid *Acyrthosiphon pisum* (Ferrari et al. 2004), the bird cherry oat aphid *Rhopalosiphum padi* (Leybourne et al. 2020), the cowpea aphid *A. craccivora* (Asplen et al. 2014) or the black bean aphid *Aphis fabae* (Schmid et al. 2012) – against parasitoid wasps. However, to fulfil its function, *H. defensa* relies on a partner in crime, a bacteriophage referred to as APSE (*A. pisum* secondary endosymbiont). APSE phages are integrated in the *H. defensa* genome and occur in variants that encode different putative toxins (Degnan and Moran 2008; Oliver and Higashi 2019). Spontaneous loss of APSE is associated with the loss of protection against parasitoids and overreplication of *H. defensa* (Oliver et al. 2009; Weldon et al. 2013).

Transmission mode influences coevolution

When symbionts transmit vertically, they tie their fate to their host's wellbeing. As their dispersal depends completely on their host's reproduction, they are expected to evolve towards lower virulence (Bull et al. 1991). Yet that does not mean it is all smooth sailing. Host and symbiont still are in a battle of constant co-adaptation (Bennett and Moran 2015; Stoy et al. 2020).

For bacterial symbionts, an important consequence of their host-bound lifestyle is a relatively low effective population size: At each host generation they are subject to bottlenecks, as a relatively small number of endosymbionts is transferred from mother to offspring (Kaltenpoth et al. 2010; Mira and Moran 2002). The symbionts therefore become vulnerable to genetic drift, which over time leads to gene loss through accumulation of deleterious mutations (Moran and Bennett 2014; Pettersson and Berg 2007). To avoid or to counteradapt to the loss of endosymbiont functions, selection at the host level is important (Moran and Bennett 2014; Pettersson and Berg 2007). An additional problem to the symbiosis is within-host selection. If a spontaneous mutation created endosymbionts with increased growth rate and cost, within-host selection would select for the spread of these selfish endosymbionts. Selection between hosts is therefore considered vital to prevent symbionts from becoming exploitative, i.e. providing benefits at increased cost to the host (Stoy et al. 2020).

Even if symbionts of aphids are very efficient at vertical transmission (Darby and Douglas 2003; Vorburger et al. 2017) they can and do transmit horizontally, at least occasionally (Henry et al. 2013; Russell et al. 2003; Sandström et al. 2001). Horizontal transmission is expected to be associated with higher virulence than vertical transmission, as symbiont

dispersal is not tied to the host's reproduction (Ewald 1983; Fisher et al. 2017; Stewart et al. 2005). Additionally, it bears the risk of escalating virulence if several symbionts compete for resources of the same host (Frank 1992; McLean et al. 2018; Nowak and May 1994). While horizontal transmission does not preclude the evolution of mutualistic symbioses (see Nyholm and McFall-Ngai 2004 for an example of symbiosis between squids and *Vibrio*), it might have less potential to facilitate the evolution of defensive symbioses than vertical transmission (Vorburger and Perlman 2018). But even in vertically transmitted defensive symbioses, occasional horizontal transmission might have positive effects; for example if it allow the symbiont to (re-)acquire functions or slow its genome decay through recombination (Stoy et al. 2020).

Understanding the cost of symbiosis

The investigation of defensive symbioses, such as the one between *A. fabae* and *H. defensa,* is slowed by the difficulty to culture many endosymbiotic bacteria, which means that useful genetic tools to disrupt the function or report the activation of genes are not available. Nevertheless, we can imagine three mechanisms that could explain both the protection and the costs of defensive endosymbionts. Firstly, endosymbionts could consume host resources that are also required by parasites, such as parasitoid wasps. This would lead to the host 'paying' for the protection it receives with some of its own resources, while the parasite struggles to survive due to limited nutrient supply (Oliver et al. 2014). Secondly, toxins produced by the endosymbiont might not only kill parasites but also cause collateral damage to the host. Toxins strongly suggested to be involved in defensive symbioses are those encoded by the APSE in case of *H. defensa* (Moran et al. 2005a; Oliver et al. 2009; Weldon et al. 2013) or ribosome-inactivating proteins in case of *Spiroplasma* (Ballinger and Perlman 2017; Garcia-Arraez et al. 2019). Thirdly, endosymbionts might prime a host's immune system, which could on one hand be costly for the host (Moret and Schmid-Hempel 2000), but might also allow swifter and more efficient attack of invading parasites.

Thesis outline

My PhD studies aimed at increasing our understanding of the origin and mechanisms behind the costs of secondary endosymbionts of aphids. It was part of a SNSF-funded Sinergia project, in which three laboratories collaborated to investigate defensive symbioses in different host species. While two laboratories investigated the protective function of *Spiroplasma* in different *Drosophila* species, our group worked on *Spiroplasma* and *H.*

defensa in aphids. My own work focused on the cost of infection with *H. defensa* in *A. fabae*. In the first chapter I used transcriptome sequencing to generate hypotheses about the mechanistic basis of *H. defensa*-mediated costs to the host. Are they a side-effect of the toxins that *H. defensa* produces to fight its host's enemies, a result of resources consumed by *H. defensa,* or immune response against toxins to fight the host's enemies? In the second chapter, I assessed the horizontal transmission success in dependence of the titre of a *H. defensa* isolate. I asked whether costly isolates with high titre can offset their cost on their current host's offspring production with a higher potential for horizontal transmission. Finally, the third chapter addresses the comparability of costs measured in experimental host-endosymbiont associations to associations tested by natural selection in the field. To this end, I compared the fitness gained from losing *H. defensa* in natural host-endosymbiont associations to the fitness lost from introducing them into naturally uninfected aphid clones.

In addition to my own independent work on *H. defensa* in *A. fabae*, my PhD research also contributed to two projects concerned with bacterial endosymbionts of the pea aphid, *A. pisum*. These projects included a large field sampling campaign in Europe to characterize the natural endosymbiont community of pea aphids and to obtain a collection of *Spiroplasma* strains for a study of their phenotypic effects on pea aphid hosts. Although the lead for these two projects was with my colleague, postdoctoral scientist Hugo Mathé-Hubert, and my PhD adviser Christoph Vorburger, these projects formed an integral part of my thesis work. I was part of the field sampling crew, I was responsible for most of the molecular analyses (e.g. multilocus sequence typing of *Spiroplasma* strains), and I participated in all experiments. The results of this work are published in two papers with Hugo Mathé-Hubert as first author and myself as second author. I include these papers as an appendix to the thesis and I briefly summarise their main findings here.

Appendix chapter 1:
Hugo Mathé-Hubert, Heidi Kaech, Pravin Ganesanandamoorthy, and Christoph Vorburger
Evolutionary costs and benefits of infection with diverse strains of *Spiroplasma* in pea aphids
(2019; Evolution 73: 1466-1481)

The endosymbiont *Spiroplasma* is known to protect against parasitoid wasps and parasitoid nematodes in *Drosophila* (Jaenike et al. 2010; Paredes et al. 2016; Xie et al. 2010). To identify the evolutionary costs and benefits of *Spiroplasma* in pea aphids, we transfected a

selection of 12 *Spiroplasma* strains from *Spiroplasma*-positive pea aphids collected during the field sampling into a common genetic background. Just as *H. defensa* in *A. fabae,* virtually all *Spiroplasma* strains curtailed the pea aphid's lifespan. While there was only limited evidence that these *Spiroplasma* strains compensate their costs with the benefit of protecting their host from the parasitoid *Aphidius ervi*, another study by our collaborator Ailsa McLean has since demonstrated that *Spiroplasma* protects the pea aphid against another important parasitoid, *Praon volucre* (McLean et al. 2020). Additionally, we characterized all *Spiroplasma* strains with a multi-locus sequence type approach. This revealed three different clades that were predictive of *Spiroplasma*'s effects on host fitness (approach described in Mathé-Hubert et al. 2019).

Appendix chapter 2:
Hugo Mathé-Hubert, Heidi Kaech, Corinne Hertaeg, John Jaenike, and Christoph Vorburger
Non-random associations of maternally transmitted symbionts in insects: The roles of drift versus biased co-transmission and selection
(2019; Molecular Ecology 28: 5330-5346)

Different defensive endosymbionts of a host species might interact in several possible ways. If two provide a similar service, one of them might be redundant and association of the symbionts in the same host might therefore be rare. Alternatively, presence of one symbiont might mitigate or exacerbate another's costs, making co-occurrence beneficial or detrimental to the host. This could lead to positive or negative associations between different symbiont species in aphid populations. However, such associations may also occur through drift alone, complicating the interpretation of co-occurrence patterns. In our collection of 498 pea aphids from across Europe (France, Switzerland, Germany and Denmark), the co-occurrence of several symbiont species was more or less common than expected under random assortment. Hugo Mathé-Hubert thus developed a model that allowed assessing the effect of drift on symbiont co-occurrence. Based on this model it was possible to conclude that some endosymbiont associations were too strong to be explained by drift, pointing at ecologically relevant interactions between symbionts. This implies that studying endosymbionts in isolation will only provide an incomplete picture of their effects on host fitness.

Literature cited

Asplen MK, Bano N, Brady CM, Desneux N, Hopper KR, Malouines C, Oliver KM, White JA, Heimpel GE. 2014. Specialisation of bacterial endosymbionts that protect aphids from parasitoids. Ecological Entomology. 39(6):736-739.

Ballinger MJ, Perlman SJ. 2017. Generality of toxins in defensive symbiosis: ribosome-inactivating proteins and defense against parasitic wasps in *Drosophila*. Plos Pathog. 13(7):e1006431.

Baumann P. 1995. Genetics, Physiology, and Evolutionary Relationships of the Genus *Buchnera*: Intracellular Symbionts of Aphids. Annual Review of Microbiology. 49:55-94.

Bennett GM, Moran NA. 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. Proceedings of the National Academy of Sciences. 112(33):10169.

Brinza L, Viñuelas J, Cottret L, Calevro F, Rahbé Y, Febvay G, Duport G, Colella S, Rabatel A, Gautier C et al. 2009. Systemic analysis of the symbiotic function of *Buchnera aphidicola*, the primary endosymbiont of the pea aphid *Acyrthosiphon pisum*. Comptes Rendus Biologies. 332(11):1034-1049.

Bull JJ, Molineux IJ, Rice WR. 1991. Selection of benevolence in a host-parasite system. Evolution. 45(4):875-882.

Cayetano L, Rothacher L, Simon J-C, Vorburger C. 2015. Cheaper is not always worse: Strongly protective isolates of a defensive symbiont are less costly to the aphid host. Proceedings of the Royal Society B. 282(1799).

Darby AC, Douglas AE. 2003. Elucidation of the transmission patterns of an insect-borne bacterium. Appl Environ Microb. 69(8):4403-4407.

Dedryver C-A, Le Ralec A, Fabre F. 2010. The conflicting relationships between aphids and men: A review of aphid damage and control strategies. Comptes Rendus Biologies. 333(6):539-553.

Degnan PH, Moran NA. 2008. Diverse phage-encoded toxins in a protective insect endosymbiont. Appl Environ Microb. 74(21):6782-6791.

Dion E, Polin SE, Simon J-C, Outreman Y. 2011. Symbiont infection affects aphid defensive behaviours. Biology letters. 7(5):743-746.

Douglas AE. 1998. Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria *Buchnera*. Annual Review of Entomology. 43(1):17-37.

Douglas AE. 2016. How multi-partner endosymbioses function. Nature Reviews Microbiology. 14(12):731-743.

Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstädter J, Hurst GD. 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. BMC Biol. 6(1):27.

Dykstra HR, Weldon SR, Martinez AJ, White JA, Hopper KR, Heimpel GE, Asplen MK, Oliver KM. 2014. Factors limiting the spread of the protective symbiont *Hamiltonella defensa* in *Aphis craccivora* aphids. Appl Environ Microb. 80(18):5818-5827.

Ewald PW. 1983. Host-parasite relations, vectors, and the evolution of disease severity. Annual Review of Ecology and Systematics. 14:465-485.

Ferrari J, Darby AC, Daniell TJ, Godfray HCJ, Douglas AE. 2004. Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. Ecological Entomology. 29(1):60-65.

Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. 2017. The evolution of host-symbiont dependence. Nature Communications. 8(1):15973.

Frago E, Mala M, Weldegergis BT, Yang C, McLean A, Godfray HCJ, Gols R, Dicke M. 2017. Symbionts protect aphids from parasitic wasps by attenuating herbivore-induced plant volatiles. Nature Communications. 8(1):1860.

Frank SA. 1992. A kin selection model for the evolution of virulence. Proceedings of the Royal Society of London Series B: Biological Sciences. 250(1329):195-197.

Fukatsu T, Nikoh N, Kawai R, Koga R. 2000. The secondary endosymbiotic bacterium of the pea aphid *Acyrthosiphon pisum* (Insecta: Homoptera). Appl Environ Microbiol. 66(7):2748-2758.

Garcia-Arraez MG, Masson F, Escobar JCP, Lemaitre B. 2019. Functional analysis of RIP toxins from the *Drosophila* endosymbiont *Spiroplasma poulsonii*. BMC Microbiology. 19(1):46.

Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. P Natl Acad Sci USA. 108(7):2849-2854.

Henry LM, Peccoud J, Simon JC, Hadfield JD, Maiden MJC, Ferrari J, Godfray HCJ. 2013. Horizontally transmitted symbionts and host colonization of ecological niches. Curr Biol. 23(17):1713-1717.

Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. 2008. How many species are infected with *Wolbachia*? - A statistical analysis of current data. Fems Microbiol Lett. 281(2):215-220.

Ives AR, Barton BT, Penczykowski RM, Harmon JP, Kim KL, Oliver K, Radeloff VC. 2020. Self-perpetuating ecological–evolutionary dynamics in an agricultural host–parasite system. Nature Ecology & Evolution. 4(5):702-711.

Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ. 2010. Adaptation via symbiosis: Recent spread of *Drosophila* defensive symbiont. Science. 329(5988):212.

Jamin AR, Vorburger C. 2019. Estimating costs of aphid resistance to parasitoids conferred by a protective strain of the bacterial endosymbiont *Regiella insecticola*. Entomol Exp Appl. 167(3):252-260.

Kaltenpoth M, Goettler W, Koehler S, Strohm E. 2010. Life cycle and population dynamics of a protective insect symbiont reveal severe bottlenecks during vertical transmission. Evolutionary Ecology. 24(2):463-477.

Leybourne DJ, Bos JIB, Valentine TA, Karley AJ. 2020. The price of protection: A defensive endosymbiont impairs nymph growth in the bird cherry-oat aphid, *Rhopalosiphum padi*. Insect science. 27(1):69-85.

Łukasik P, van Asch M, Guo H, Ferrari J, Charles J. Godfray H. 2013. Unrelated facultative endosymbionts protect aphids against a fungal pathogen. Ecology Letters. 16(2):214-218.

Mathé-Hubert H, Kaech H, Hertaeg C, Jaenike J, Vorburger C. 2019. Nonrandom associations of maternally transmitted symbionts in insects: The roles of drift versus biased cotransmission and selection. Molecular Ecology. 28(24):5330-5346.

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF et al. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proceedings of the National Academy of Sciences. 110(9):3229.

McLean AHC, Hrček J, Parker BJ, Mathé-Hubert H, Kaech H, Paine C, Godfray HCJ. 2020. Multiple phenotypes conferred by a single insect symbiont are independent. Proceedings of the Royal Society B: Biological Sciences. 287(1929):20200562.

McLean AHC, Parker BJ, Hrček J, Kavanagh JC, Wellham PAD, Godfray HCJ. 2018. Consequences of symbiont co-infections for insect host phenotypes. Journal of Animal Ecology. 87(2):478-488.

Mira A, Moran NA. 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microbial Ecology. 44(2):137-143.

Montllor CB, Maxmen A, Purcell AH. 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. Ecological Entomology. 27(2):189-195.

Moran N, Baumann P. 1994. Phylogenetics of cytoplasmically inherited microorganisms of arthropods. Trends in Ecology & Evolution. 9(1):15-20.

Moran NA, Bennett GM. 2014. The tiniest tiny genomes. Annual Review of Microbiology. 68(1):195-215.

Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H. 2005a. The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes. P Natl Acad Sci USA. 102(47):16919-16926.

Moran NA, Russell JA, Koga R, Fukatsu T. 2005b. Evolutionary relationships of three new species of *Enterobacteriaceae* living as symbionts of aphids and other insects. Appl Environ Microb. 71(6):3302-3310.

Moret Y, Schmid-Hempel P. 2000. Survival for immunity: The price of immune system activation for bumblebee workers. Science. 290(5494):1166.

Nowak MA, May RM. 1994. Superinfection and the evolution of parasite virulence. Proceedings of the Royal Society of London Series B: Biological Sciences. 255(1342):81-89.

Nyholm SV, McFall-Ngai M. 2004. The winnowing: Establishing the squid–*Vibrio* symbiosis. Nature Reviews Microbiology. 2(8):632-642.

Oliver KM, Campos J, Moran NA, Hunter MS. 2008. Population dynamics of defensive symbionts in aphids. Proceedings of the Royal Society B. 275(1632):293-299.
Oliver KM, Degnan PH, Hunter MS, Moran NA. 2009. Bacteriophages Encode Factors Required for Protection in a Symbiotic Mutualism. Science. 325(5943):992-994.
Oliver KM, Higashi CH. 2019. Variations on a protective theme: *Hamiltonella defensa* infections in aphids variably impact parasitoid success. Current opinion in insect science. 32:1-7.
Oliver KM, Smith AH, Russell JA. 2014. Defensive symbiosis in the real world – advancing ecological studies of heritable, protective bacteria in aphids and beyond. Functional Ecology. 28(2):341-355.
Paredes JC, Herren JK, Schüpfer F, Lemaitre B. 2016. The role of lipid competition for endosymbiont-mediated protection against parasitoid asps in *Drosophila*. Mbio. 7(4):e01006-01016.
Pettersson ME, Berg OG. 2007. Muller's ratchet in symbiont populations. Genetica. 130(2):199-211.
Polin S, Simon J-C, Outreman Y. 2014. An ecological cost associated with protective symbionts of aphids. Ecology and Evolution. 4(6):836-840.
Rossbacher S, Vorburger C. 2020. Prior adaptation of parasitoids improves biological control of symbiont-protected pests. Evolutionary Applications. 13(8):1868-1876.
Russell JA, Latorre A, Sabater-Muñoz B, Moya A, Moran NA. 2003. Side-stepping secondary symbionts: Widespread horizontal transfer across and beyond the Aphidoidea. Mol Ecol. 12(4):1061-1075.
Sandström JP, Russell JA, White JP, Moran NA. 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. Mol Ecol. 10(1):217-228.
Schmid M, Sieber R, Zimmermann Y-S, Vorburger C. 2012. Development, specificity and sublethal effects of symbiont-conferred resistance to parasitoids in aphids. Functional Ecology. 26(1):207-215.
Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature. 407(6800):81-86.
Smith AH, Łukasik P, O'Connor MP, Lee A, Mayo G, Drott MT, Doll S, Tuttle R, Disciullo RA, Messina A et al. 2015. Patterns, causes and consequences of defensive microbiome dynamics across multiple scales. Molecular Ecology. 24(5):1135-1149.
Smith TE, Moran NA. 2020. Coordination of host and symbiont gene expression reveals a metabolic tug-of-war between aphids and *Buchnera*. Proceedings of the National Academy of Sciences. 117(4):2113.
Stewart AD, Logsdon JM, Kelley SE. 2005. An empirical study of the evolution of virulence under both horizontal and vertical transmission. Evolution. 59(4):730-739.
Stoy KS, Gibson AK, Gerardo NM, Morran LT. 2020. A need to consider the evolutionary genetics of host–symbiont mutualisms. Journal of Evolutionary Biology. 00:1-13.
Van Emden HF, Harrington R. 2007. Aphids as crop pests. Wallingford: CABI.
Vorburger C, Perlman SJ. 2018. The role of defensive symbionts in host–parasite coevolution. Biological Reviews. 93(4):1747-1764.
Vorburger C, Rouchet R. 2016. Are aphid parasitoids locally adapted to the prevalence of defensive symbionts in their hosts? Bmc Evol Biol. 16(1):271.
Vorburger C, Siegrist G, Rhyner N. 2017. Faithful vertical transmission but ineffective horizontal transmission of bacterial endosymbionts during sexual reproduction of the black bean aphid, *Aphis fabae*. Ecological Entomology. 42(2):202-209.
Weldon SR, Strand MR, Oliver KM. 2013. Phage loss and the breakdown of a defensive symbiosis in aphids. P Roy Soc B-Biol Sci. 280(1751).
Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: Master manipulators of invertebrate biology. Nature Reviews Microbiology. 6(10):741-751.
Wilkinson TL, Fukatsu T, Ishikawa H. 2003. Transmission of symbiotic bacteria *Buchnera* to parthenogenetic embryos in the aphid *Acyrthosiphon pisum* (Hemiptera: Aphidoidea). Arthropod Structure and Development. 32(2-3):241-245.
Xie J, Vilchez I, Mateos M. 2010. *Spiroplasma* bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma*. PLOS ONE. 5(8):e12149.

Chapter 1

Triple RNA-Seq characterizes aphid gene expression in response to infection with unequally virulent strains of the endosymbiont *Hamiltonella defensa*

Heidi Kaech[1,2], Alice B. Dennis[3], Christoph Vorburger[1,2,*]

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[2]D-USYS, Department of Environmental Systems Science, ETH Zürich, Switzerland
[3]Institute of Biochemistry and Biology, University Potsdam, Germany

Publication in preparation

Abstract

Secondary endosymbionts of aphids provide benefits to their hosts, but also impose costs such as reduced lifespan and reproductive output. The aphid *Aphis fabae* is host to different strains of the secondary endosymbiont *Hamiltonella defensa*, which encode different putative toxins. These strains have very different phenotypes: They reach different densities in the host, and the costs and benefits (protection against parasitoid wasps) they confer to the host vary strongly. We used RNA-Seq to generate hypotheses on why four of these strains inflict such different costs to *A. fabae*. We found different *H. defensa* strains to cause strain-specific changes in aphid gene expression, but little effect of *H. defensa* on gene expression of the primary endosymbiont, *Buchnera aphidicola*. The most costly *H. defensa* strain, H85, was associated with strongly reduced aphid expression of hemocytin, a marker of insect hemocytes. The downregulation of hemocytin could represent a loss of control of the host over the secondary endosymbiont, which would explain the high density that strain H85 reaches in the host. Overall, our results suggest that costs of different strains of *H. defensa* are likely caused by different mechanisms, and that these costs are imposed by interacting with the host rather than the host's obligatory endosymbiont *B. aphidicola*.

Keywords: *Aphis fabae, Buchnera*, cost of resistance, *Hamiltonella*, host-symbiont interaction, RNA-Seq, symbiosis

Introduction

Insects have a complicated relationship with bacteria. On one hand, they are exposed to environmental bacterial pathogens, against which their immune system should defend them (Sanchez-Contreras and Vlisidou 2008). On the other hand, insects commonly harbour beneficial bacterial endosymbionts, which their immune system should tolerate (Ratzka et al. 2012). In aphids, tolerance of the primary bacterial endosymbiont *Buchnera aphidicola* is necessary for survival, as *B. aphidicola* supplements the aphids' protein-poor diet with essential amino acids (Brinza et al. 2009; Douglas 1998; Hansen and Moran 2011; Moran and Baumann 1994). This ancient symbiosis, which is at least 160 Million years old (Moran and Baumann 1994), may be facilitated by the confinement of *B. aphidicola* to the intracellular space, where specialised bacteriocytes provide a safe space for the host to keep the bacteria (Ratzka et al. 2012). *Buchnera aphidicola* is vertically transmitted from mother to offspring (Koga et al. 2012).

Aphids also maintain a range of secondary bacterial endosymbionts. Like *B. aphidicola*, these secondary endosymbionts provide benefits, are vertically transmitted, and some of them can be found intracellularly (Moran et al. 2005b; Oliver et al. 2010). Unlike *B. aphidicola*, however, they are not strictly required for survival and also colonise the extracellular space (Moran et al. 2005b). In fact, their density in the hemolymph is sufficiently high to allow horizontal transmission to other aphids, both via artificial microinjection of hemolymph or naturally via vectors such as parasitoid wasps (Gehrer and Vorburger 2012).

The continuous presence of secondary endosymbionts in the hemolymph suggests that the aphids' immune system allows their presence. Maintenance of secondary endosymbionts might partially be attributable to peculiarities of the aphids' immune system. Comparative genomics of *Drosophila melanogaster* and the pea aphid, *Acyrthosiphon pisum,* suggest a reduced immune system repertoire in the latter. In the pea aphid, one of the two humoral response pathways, the immune deficiency (IMD) pathway, which is preferentially activated by Gram-negative bacteria in *Drosophila* (Lemaitre and Hoffmann 2007), lacks several key proteins and pattern recognition receptors (Gerardo et al. 2010). It was proposed that this facilitated the association of aphids with their mostly Gram-negative endosymbionts (Guo et al. 2017; Laughton et al. 2011). In support of this, pea aphids react strongly to heat-killed fungi, but only weakly to heat-killed Gram-negative pathogens (Barribeau et al. 2014; Gerardo et al. 2010), and experimental infection with Gram-negative *Escherichia coli* is fatal

to pea aphids (Altincicek et al. 2011). The immune response to Gram-negative bacteria may be inefficient in aphids, but it is not non-existent; in response to infection with *Serratia marcescens,* pea aphids mount a seemingly IMD-independent activation of the c-Jun N-terminal kinase (JNK) pathway (Renoz et al. 2015) and upon challenge with *E. coli,* hemocytes readily destroy *E. coli* through phagocytosis (Laughton et al. 2011; Schmitz et al. 2012). Secondary symbionts might have to protect themselves from these remnant immune responses to allow stable association with their host.

If secondary endosymbionts need to subvert the aphids' immune system to be maintained, the question arises whether aphids can control the endosymbiont density at all. Lack of control could be dangerous to the host. On one hand, secondary endosymbionts provide benefits, such as defence against pathogens (Scarborough et al. 2005), protection from parasitoids (Oliver et al. 2005), adaptation to host plants (Wagner et al. 2015), or heat shock tolerance (Russell and Moran 2006). On the other hand, secondary endosymbionts only occur at intermediate frequencies in aphid populations (Smith et al. 2015; Vorburger and Rouchet 2016). Their spread through the host populations appears to be constrained by costs, which are apparent when populations of the same aphid genotype with and without secondary endosymbionts compete against each other in experimental populations (Dykstra et al. 2014; Hafer-Hahmann and Vorburger 2020; Oliver et al. 2008). If secondary endosymbionts are inherently costly, the host should profit from controlling their density so that the optimal balance between their costs and benefits is achieved. Whether such control exists in aphids and how it might be achieved is yet unknown.

A frequent secondary endosymbiont of aphids is *Hamiltonella defensa*. It provides protection against aphid parasitoids such as *Aphidius ervi* (Oliver et al. 2003) and *Lysiphlebus fabarum* (Schmid et al. 2012; Vorburger et al. 2009). Although *H. defensa* itself encodes putative toxins, which could potentially hinder parasitoid development, the strongest link to its protective function is the lysogenic bacteriophage APSE (*A. pisum* secondary endosymbiont) (Moran et al. 2005a; Oliver et al. 2009). This phage is integrated in the *H. defensa* genome and occurs in variants that encode different putative toxins (Degnan and Moran 2008; Oliver and Higashi 2019). Spontaneous loss of APSE in strains hosted by pea aphids is associated with the loss of protection against parasitoids and over-replication of *H. defensa* (Oliver et al. 2009; Weldon et al. 2013). In the black bean aphid (*Aphis fabae*), *H. defensa* and its associated APSE lead to a reduced lifespan and lifetime reproduction in the absence of parasitoids (Vorburger and Gouskov 2011). Possible explanations include the resource

consumption by the endosymbiont population, collateral damage to the host from the APSE's toxins, or the energy requirements of immune activation if secondary endosymbionts have to be controlled by the aphid's immune system (Vorburger and Perlman 2018).

We would expect to see a positive correlation between cost and benefit in natural populations of secondary endosymbionts: The more beneficial an endosymbiont is to its host, the more costly it can be and still spread in the host population. For *H. defensa* in black bean aphids, a comparison of 11 strains suggested that some strains strongly protect hosts from parasitation by *L. fabarum* but have little impact on host longevity and offspring production, while others are more weakly protective but highly costly (Cayetano et al. 2015 and Supplementary Figure 1). Thus, cost and benefit of different *H. defensa* strains are not connected in such a straightforward way as expected.

In this work, we investigate four *H. defensa* phenotypes – ranging from the apparently mutualistic strain H76 to the over-replicating and pathogen-like strain H85. Using 'triple' RNA-sequencing to measure gene expression of *A. fabae* hosts, their obligate endosymbiont *B. aphidicola* and their secondary endosymbiont *H. defensa,* we aim to generate hypotheses about how *H. defensa* inflicts costs on the black bean aphid host and whether the host regulates the density of *H. defensa*.

Results

Sequencing output

We sequenced the transcriptome of aphids carrying only their obligatory endosymbiont *B. aphidicola* (H0) and identically reared aphids from the same genetic background that previously were experimentally infected by one out of four different *H. defensa* strains: H15, H76, H85 or H402. Each of the five treatment was replicated four times (R1-R4) and each replicate consisted of a pool of 18 aphids. In total, two sequencing runs yielded 755.9 million read pairs. Strict trimming for *de novo* assembly retained 83% of the read pairs. Relaxed trimming for mapping retained 93% of the read pairs, or an average of 34.7 million (± 2.8 SD) read pairs per library (Supplementary Table 1). Reads trimmed for *de novo* assembly and mapping were both quality screened by FastQC (Andrews 2010).

All libraries were checked for contamination. One of the 20 libraries, library H15R1, was heavily contaminated with reads of human and human-associated bacterial origin (Supplementary Table 2). This library also took an outlier position in a PCA that segregated

libraries based on aphid gene expression patterns (Supplementary Figure 2) and was therefore excluded from further analyses.

Assembly

Our approach could be called a 'triple' RNA-Seq because it contains transcripts from three organisms – aphid host, obligatory endosymbiont and secondary endosymbiont. These were assembled *de novo* and assigned to the most likely organism of origin using blast. Chimeric transcripts were discarded upon detection and prokaryote multi-gene transcripts were split to allow gene annotation and differential gene expression analysis.

For aphids, the assembly generated 46'352 transcripts. Transcript length ranged from 297 to 27'541 nucleotides (mean length: 2'657.9 bp, N_{50}: 3'542 bp, GC: 32.02%). 1'255 transcripts had no blast results and were discarded from analysis. The remaining transcripts were assigned to a total of 10'809 genes, of which 7'313 could be annotated with GO-terms using OmicsBox. In comparison, the genome of *Aphis glycines* contains 17'558 genes (Wenger et al. 2017). In our assembly, 93.1% of the *insecta* BUSCO genes were complete, while 2.6% were fragmented (Supplementary Table 3).

Transcripts blasting to *B. aphidicola* were aligned to the reference genomes from *B. aphidicola* isolated from *A. glycines* and *B. aphidicola* from *Uroleucon ambrosiae.* This produced 616 genes with a GC content of 25.2% and an N_{50} of 1'206 bases. Of these genes, 569 could be annotated with GO-terms. In comparison, *B. aphidicola* of *A. glycines* has 618 genes. Our assembly reached a *proteobacteria* BUSCO score of 73.3% complete genes (Supplementary Table 3). Such a low value was expected due to the reduced genome of *B. aphidicola.*

Transcriptomes of different *H. defensa* strains were assembled from libraries containing the strain. From the four assemblies, 4'850 transcripts were retrieved. From these transcripts we identified 1'706 *H. defensa* and APSE genes. GC content of the genes was 41.35% and 1'326 genes could be annotated with GO-terms. In comparison, *H. defensa* strain ZA17, from *A. pisum*, contains 2'370 genes. In our assembly, 92.3% of the *proteobacteria* BUSCO genes were complete, 3.2% were fragmented (Supplementary Table 3).

Mapping

Over all 19 libraries included in the analysis, 73% of read pairs could be mapped (Supplementary Table 1). Across all libraries, the majority of read pairs (61%) mapped to

aphid genes. To *B. aphidicola* genes, 8.2% of the reads were mapped. The ratio of *B. aphidicola* to aphid reads was stable across treatments (Figure 1 A). In contrast, the ratio of *H. defensa* to aphid reads was much higher in aphids infected with *H. defensa* H85 than in aphids infected by other strains (Figure 1 B). For H85, the average percentage of mapped reads was 12.7%. For H76 and H402, intermediate numbers of read pairs were recorded (1.4 % and 1.5%, respectively). The percentage of mapped read pairs was lowest for H15 (0.6%). The APSE to *H. defensa* read pair ratio was highest in H76, intermediate in H402 and lowest in H15 and H85 (Figure 1 C).

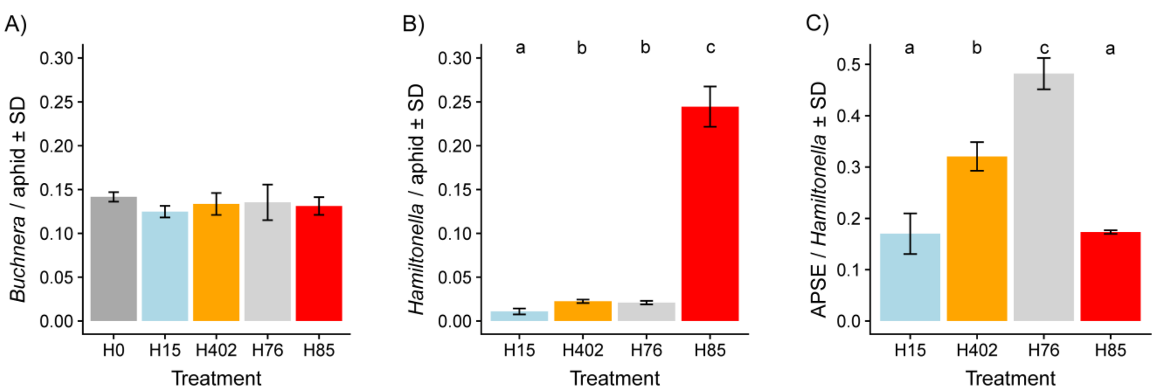


Figure 1 – Overreplication of *H. defensa* strain H85. A) Ratio of reads mapped to *B. aphidicola* and to aphid genes, averaged by treatment (uninfected (H0, dark grey) or *H. defensa*-infected aphid hosts (infecting strains H15 (blue), H402 (orange), H76 (light grey), H85 (red)). A one-way ANOVA comparing the effect of treatment on the read ratio was not significant ($F_{(4,14)}$= 0.84, p=0.52). B) Ratio of reads mapped to *H. defensa* genes and to aphid genes. A one-way ANOVA comparing the effect of treatment on the log read ratio was significant ($F_{(3,11)}$= 275.57, p<0.001). Treatments with different letters are significantly different in pairwise post-hoc tests (Tukey's HSD). (C) Ratio of reads mapped to APSE genes and to *H. defensa* genes. A one-way ANOVA comparing the effect of treatment on the read ratio was significant ($F_{(3,11)}$= 109.77, p<0.001). Treatments with different letters are significantly different in pairwise post-hoc tests (Tukey's HSD).

Even though aphids from treatment H0 were not infected by *H. defensa,* a small number of read pairs were assigned to *H. defensa* genes in H0 libraries. This likely was a result of index hopping (see Methods), and it allowed us to estimate that approx. 0.2% of read pairs had undergone index hopping, which falls just below the range of 0.3-0.5% expected by the manufacturer (Illumina Inc., 2018). Given evidence that all possible index hopping combinations occur at uniform distribution around the mean (Costello et al. 2018), the influence of index hopping on fold change values was considered negligible.

Differential gene expression in aphids

For differential gene expression analysis, lowly expressed genes were removed from analysis, leaving 8'614 genes. Gene expression of aphids infected by each of the four *H. defensa* strains was individually compared to gene expression of uninfected aphids (H0) (Figure 2 A). There

were between 11 and 42 differentially expressed genes (DEG) (Figure 2 B, Supplementary Table 4). Out of the 81 aphid genes affected by the presence of *H. defensa*, only three were differentially expressed in the presence of all four *H. defensa* strains: G patch domain-containing protein 11, an uncharacterized protein and peptide chain release factor 1 (Figure 2 B, Supplementary Table 5).

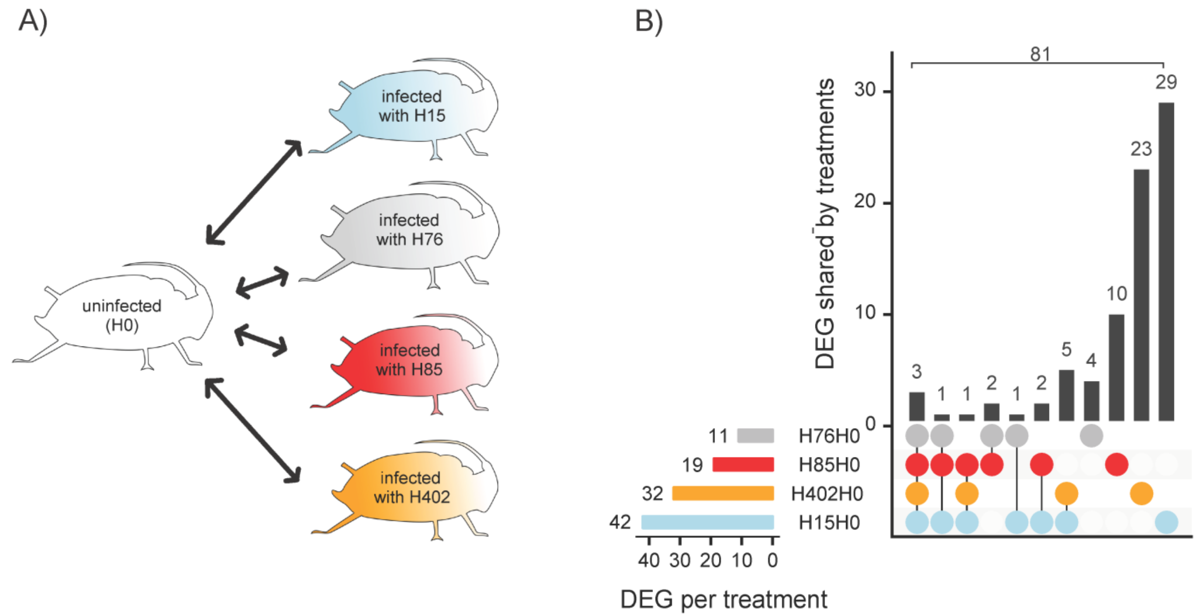


Figure 2 – Few differentially expressed aphid genes between treatments. (A) Gene expression is compared between aphids infected by *H. defensa* (infecting strains: H15 (blue), H402 (orange), H76 (grey) or H85 (red)) and uninfected aphids (H0). (B) The horizontal bars indicate the total number of differentially expressed genes (DEG) per treatment. Vertical bars indicate which genes are differentially expressed in all four treatments (leftmost column), in two or three treatments (middle columns) or in only one treatment (rightmost four columns). The sum of all vertical bars corresponds to the total number of affected genes over all four treatments.

The most prominent changes to gene expression were observed between aphids infected with H402 and uninfected aphids. In a PCA of aphid gene expression patterns, libraries of treatment H402 were clearly separated from other treatments (Figure 3), and the median log2 fold change of the 32 DEG between H402 and H0 was higher than when aphids were infected by other *H. defensa* strains (Supplementary Table 4). The function of 25 of the 32 DEG could not be determined; blasting against nucleotide and protein databases only yielded references to uncharacterized proteins. Libraries of treatments other than H402 clustered closer to the control treatment H0, which was also reflected in lower median fold changes (Supplementary Table 4). Aphids infected with H15 differentially expressed 42 genes compared to H0, aphids infected with H85 differentially expressed 19 genes and aphids infected with H76 differentially expressed 11 genes compared to H0.

To retrieve a functional profile of the DEG, we tested for GO-term enrichment. However, no GO-terms were found enriched after correcting for multiple testing, regardless of whether we analysed DEG of each treatment or DEG shared between different treatments.

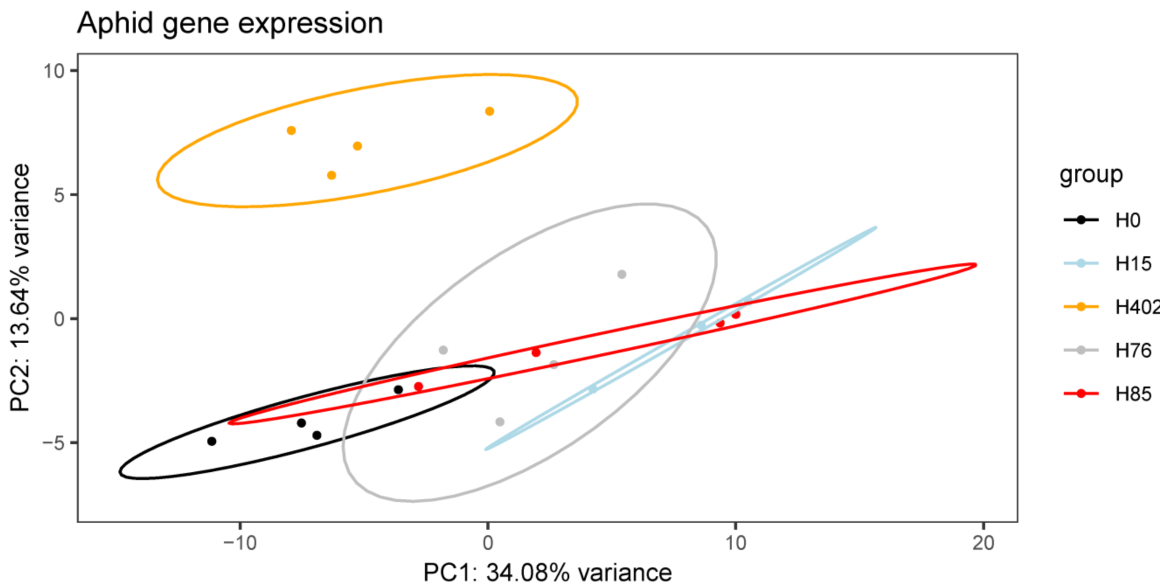


Figure 3 – Aphid gene expression changes most upon infection with *H. defensa* strain H402. PCA of the normalised and variance stabilisation transformed read count of all aphid genes expressed in uninfected (H0, black) and *H. defensa* infected aphids (infecting strains: H15 (blue), H402 (orange), H76 (grey), H85 (red)).

To investigate the difference in aphid phenotype caused by the genotypically similar *H. defensa* strains H15 and H85, we also compared aphids infected by H85 to aphids infected by H15 (Supplementary Table 6). In this comparison, only six genes were differentially expressed between H85 and H15. Aphids infected with H85 upregulated three genes (protein aubergine, nuclear pore complex protein Nup50 and ubiquitin-related modifier 1) and downregulated two uncharacterized proteins as well as hemocytin. Aphids infected by H85 express less hemocytin than aphids infected by H15 as well aphids infected by H76 or H402 and uninfected aphids (Figure 4 A). Hemocytin is linked to the insect immune system: The homolog of hemocytin in *Drosophila melanogaster* is known as hemolectin (*hml*), and genes of the *hml* family are marker of hemocytes (Goto et al. 2001). Other hemocyte markers detected in our gene expression data – croquemort (*crq*), protein singed (*sn*), protein lozenge (*lz*) and two transcripts annotated as peroxidasin (*pxn*) (Franc et al. 1996; Goto et al. 2001; Lebestky et al. 2000; Nelson et al. 1994; Zanet et al. 2009) – were not significantly differentially expressed (Figure 4 B-E).

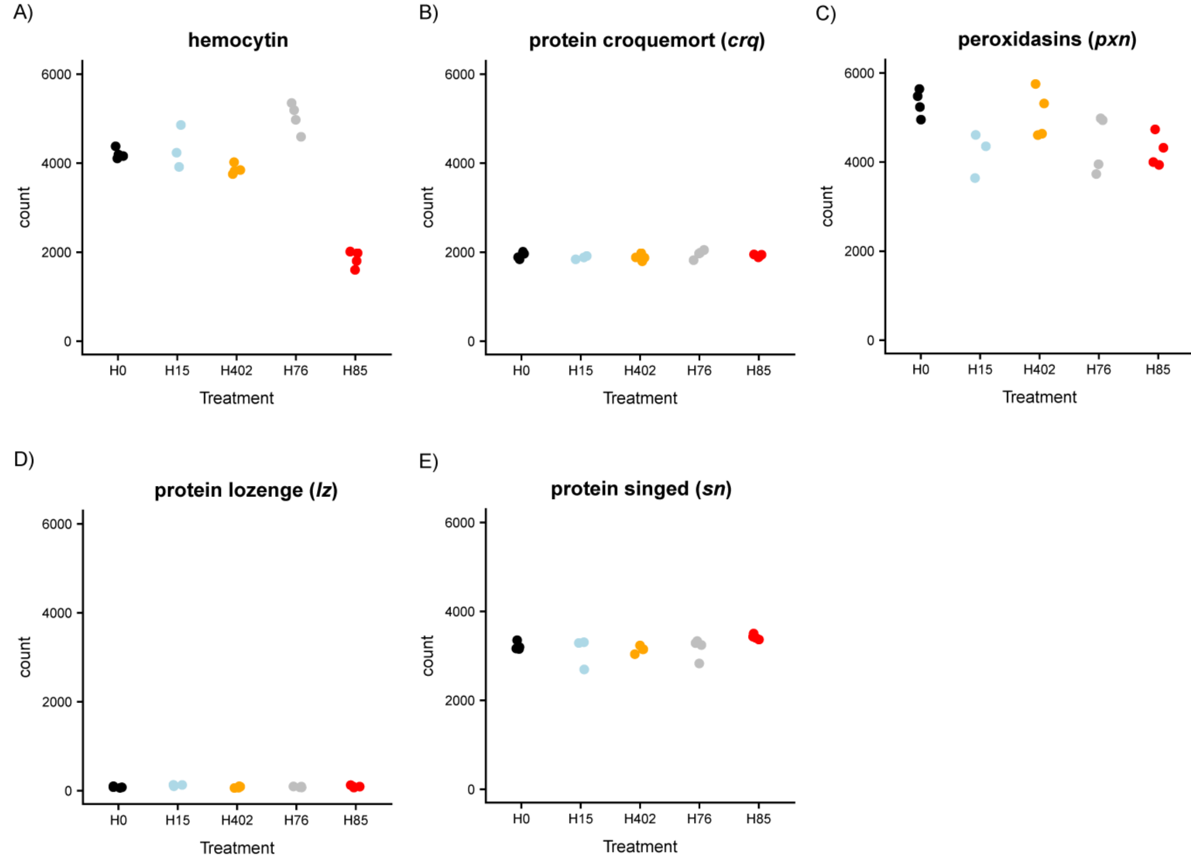


Figure 4 – One of the aphid's hemocyte marker genes is downregulated in presence of H. *defensa* strain H85. Normalised read counts of (A) hemocytin, (B) protein croquemort, (D) protein lozenge and (E) protein singed. For (C) read counts of two transcripts annotated with "peroxidasin" and "Low quality protein: peroxidasin" were combined. Aphids were either infected by *H. defensa* (strains H15 (blue), H402 (orange), H76 (grey) or H85 (red)) or uninfected (H0, black).

Differential gene expression between *Hamiltonella defensa* strains

For all analyses, gene expression of *H. defensa* and their APSE bacteriophage was combined and will be referred to as "*H. defensa* gene expression". Low-expression genes were removed from the analyses, leaving 1'477 genes. A PCA of *H. defensa* gene expression patterns segregated H76 and H402 distinctly from H15 and H85 (Figure 5 A). As with aphid expression, we conducted a separate analysis comparing just H15 and H85; this showed a clear distinction in gene expression patterns between these two strains as well (Figure 5 B).

To assess differences between the four *H. defensa* strains, we used the costly H85 as a reference. In the full model, H15 differentially expressed only 60 (or 4.1%) of 1'477 *H. defensa* genes that were included in the analysis compared to H85, but H402 and H76 differentially expressed 669 and 578 (or 46% and 39%) of all genes. To check whether the full model was destabilised by these high numbers of DEG, we also used a reduced model to

compare only strains H15 and H85. Combined, the full and reduced models found 64 DEG between H15 and H85. The majority of these, 76.6%, were reported by both models. A total of 11 genes and 4 genes were only reported by the full and by the reduced model, respectively. Based on the small differences between the two models the full model was considered stable and its results were used for further analysis.

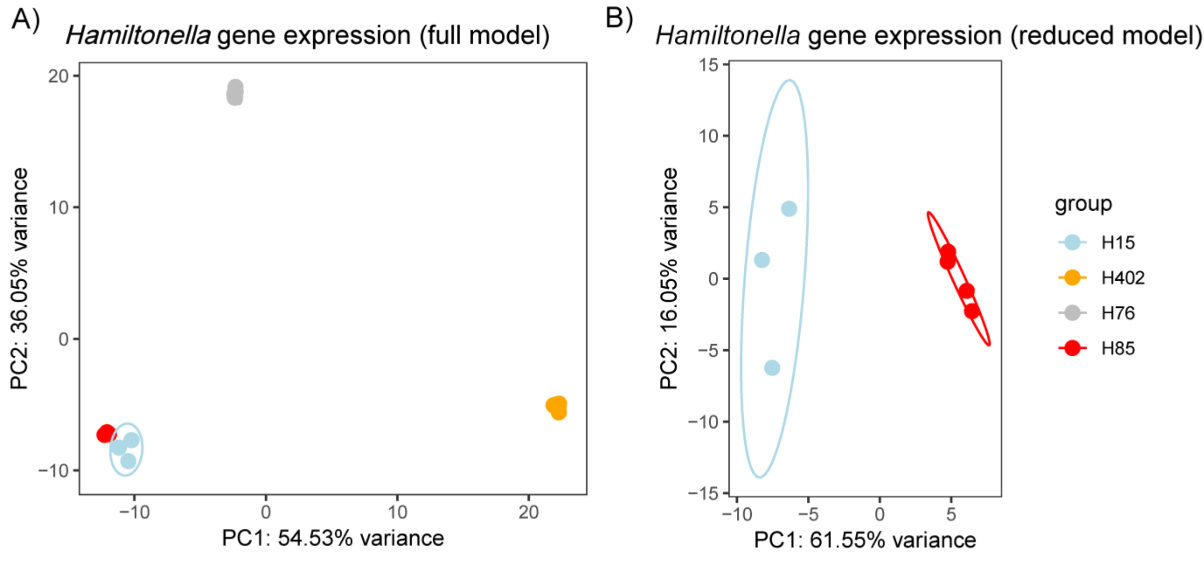


Figure 5 – Gene expression of the four *H. defensa* strains is very different. PCA of the normalised and variance stabilisation transformed read count of all genes expressed *H. defensa* (H15 (blue), H402 (orange), H76 (grey), H85 (red)). (A) Full model containing all libraries except H15R1. (B) Reduced model containing only libraries from treatment H15 and H85.The 95% confidence ellipse is sometimes covered by the dots indicating the samples' location in the PCA plot.

In the DEG between different *H. defensa* strains, seven GO-terms were significantly enriched (Table 1): 'Pathogenesis' in the DEG between H402 and H85, and GO-terms linked to translation ('structural constituent of ribosome', 'ribosome', 'rRNA binding' and 'translation') in the DEG between H15 and H85.

A total of 21 genes were differentially regulated in all three strains H15, H76 and H402 compared to H85 (Figure 6, Supplementary Table 7). These genes were not significantly enriched for any GO-terms. Strains H76 and H402 shared more than half of the genes that they differentially expressed compared to H85: 64.7% and 55.9%, respectively. The 374 shared DEG were significantly enriched for the GO-term 'interspecies interaction between organisms' (Table 1). Among the 25 genes annotated with 'interspecies interaction', 12 genes also belonged to the GO-term 'viral entry into host cells'.

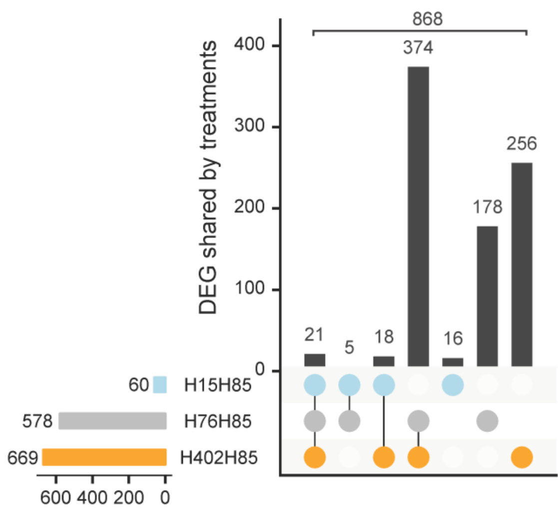


Figure 6 – Few differentially expressed *H. defensa* genes are shared between the strains, relative to H85. Gene expression of *H. defensa* strains H15 (blue), H402 (orange) and H76 (grey) in comparison to strain H85. The horizontal bars indicate the total number of differentially expressed genes (DEG) per treatment. Vertical bars indicate which genes are differentially expressed in all three treatments (leftmost column), in three or two treatments (middle columns) or in only one treatment (rightmost three columns). Data for all comparisons are from the full model with all strains. The sum of all vertical bars corresponds to the total number of affected genes over all four treatments

Table 1 – Differentially expressed Gene Ontology terms in *H. defensa.* Lists of differentially expressed *H. defensa* genes were tested for GO-term enrichment using Blast2Go's Enrichment Analysis pipeline. Lists of GO-terms were reduced to the most specific terms. GO-category, p-value and false discovery rate (FDR) are indicated for each term.

DEG List	GO-term	GO Category	p-value	FDR value
H15 vs H85	structural constituent of ribosome	Molecular function	3.99E-12	7.09E-09
	ribosome	Cellular component	1.02E-11	7.09E-09
	rRNA binding	Molecular function	3.40E-09	9.01E-07
	translation	Biological process	9.35E-08	2.07E-05
H76 vs H85	host cell membrane	Cellular component	1.40E-04	0.07
H402 vs H85	pathogenesis	Biological process	9.79E-06	0.02
Shared DEG H76 vs H85 H402 vs H85	interspecies interaction between organisms	Biological process	4.45E-06	0.01

Apart from YD-repeat toxin and CdtB toxin, we identified 31 APSE genes. All 31 APSE genes were upregulated in H76 compared to H85, while 18 were upregulated in H402 compared to H85. Between H15 and H85 no APSE genes were differentially expressed (Supplementary Table 7).

Differential gene expression in *B. aphidicola*

Based on previous studies, changes in gene expression of the obligate endosymbiont *B. aphidicola,* were expected to be subtle (Smith and Moran 2020). Indeed, of the 553 genes included in the analysis after removal of genes with low expression, only three were differentially expressed when the host was infected with *H. defensa.* One gene, a signal peptidase II showed strong variation between replicates of different treatments, leading to exclusion from analysis. The two other genes, the tRNA-threonylcarbamoyltransferase complex dimerization subunit type 1 *TsaB* and the DNA-binding transcriptional regulator *Fis* were both downregulated in presence of *H. defensa* H85 (Supplementary Table 8).

Correlation of aphid and secondary endosymbiont gene expression

To correlate gene expression between different organisms, we followed the two approaches described in Smith and Moran (2020). First, we used the correlation approach (Smith and Moran 2020), for which invariant *H. defensa* and aphid genes were removed from the data (Table 2). The rld-transformed read counts of the remaining 1’242 *H. defensa* genes and 1’288 aphid genes were correlated in all possible pairwise combinations to each other. This led to the identification of clusters of genes – referred to as modules – with similar expression patterns over all libraries of treatments H15, H76, H85 and H402. The eigengenes of the resulting 11 aphid and 13 *H. defensa* modules were correlated to detect instances where aphid and *H. defensa* gene expression correlates and thus in which the two species might influence each other’s gene expression. Note that modules were labelled with names indicating which species’ gene expression was compared (‘ApHdef’ for the comparison between aphid and *H. defensa*) and whether the module consists of aphid genes (A1-A11) or *H. defensa* genes (H1-H13).

Table 2 – Modules of co-expressed genes correlate with each other and with *H. defensa* titre. Genes were clustered according to their expression patterns across all libraries containing *H. defensa* (except the heavily contaminated library H15R1) using two approaches, a correlation approach as described in Smith and Moran (2020) and Weighted Correlation Network Analysis (WGNA). The genes used for analysis were clustered into modules of co-expressed genes. These modules were tested for GO-term enrichment, for correlation with modules of another organism and for correlation with *H. defensa* titre.

Approach	Compared organisms	Number of genes in analysis	Number of modules	Number of modules enriched for GO-terms	Modules correlating with $p<0.01$ with modules of other organism in analysis	Modules correlating with $p<0.01$ with *H. defensa* titre
Correlation	aphid	1'288	11	6	8	
	H. defensa	1'242	13	7	13	
Correlation	*H. defensa*	1'242	11	8	5	
	B. aphidicola	168	5	2	3	
WGCNA	aphid	8'600	18	11		2
	H. defensa	1'477	12	7		2
	B. aphidicola	554	7	6		1

The correlation approach identified two aphid modules that contained genes identified as interesting during the differential expression analysis. One of them was the aphid module ApHdef-A3, in which GO-term ‘ligase activity’ was enriched (Supplementary Table 9). Among the 22 genes in this module was hemocytin, a gene that was shown to be strongly downregulated in the presence of H85 by the differential gene expression analysis. The genes in ApHdef-A3 might be influenced in their expression by the genes in the *H. defensa* module

ApHdef-H10, since the eigengene of the aphid module ApHdef-A3 showed a strong negative correlation with the eigengene of the *H. defensa* module ApHdef-H10 ($r(13)=-0.90$, $p<0.001$) (Supplementary Figure 3, Supplementary Table 10). In the *H. defensa* module ApHdef-H10, no GO-terms were enriched, but the module contained the gene AS3p2_hypothetical_protein_CDS_BJP42_RS11500. This gene was found to be strongly upregulated in *H. defensa* H85 compared to all other strains in the differential gene expression analysis.

Of further interest was the aphid module ApHdef-A2, which contained – among its 141 genes – 20 of the 25 genes encoding uncharacterized aphid proteins that were strongly differentially expressed in presence of H402. The eigengene of the aphid module ApHdef-A2 correlated well with three *H. defensa* modules: ApHdef-H7 ($r(13)=0.77$, $p<0.001$) which was enriched for GO-terms associated to ATP synthesis, ApHdef-H12 ($r(13)=-0.82$, $p<0.001$) without associated GO-terms and ApHdef-H13 $r(13)=-0.81$, $p<0.001$) which was enriched for GO-terms such as 'integral component of membrane' and 'outer membrane' (Supplementary Figure 3, Supplementary Table 10).

Several additional aphid and *H. defensa* modules were conspicuous as they correlated very strongly with each other. For example, there was a strong negative correlation between the aphid module ApHdef-A1 and the *H. defensa* modules ApHdef-H5 ($r(13)=-0.9$, $p<0.001$, Supplementary Figure 3) and ApHdef-H8 ($r(13)=-0.89$, $p<0.001$, Supplementary Figure 3). Of the three modules, only module ApHdef-H8 was associated with GO-terms ('mismatch repair complex', 'outer membrane', 'DNA binding'). Finally, there was strong correlation between the aphid module ApHdef-A4, which was enriched for GO-terms related to protein folding and gene expression, and two *H. defensa* modules: module ApHdef-H11 ($r(13)=-0.91$, $p<0.001$), in which no GO-terms were enriched, and module ApHdef-H6 ($r(13)=0.92$, $p<0.001$), in which the terms 'modification of morphology or physiology of other organism involved in symbiotic interaction', 'dicarboxylic acid biosynthesis process' and 'RNA-dependent DNA biosynthetic process/polymerase activity' were enriched (Supplementary Figure 3, Supplementary Table 10). Notably, the *H. defensa* module ApHdef-H6 contained genes that were more or mainly expressed by strain H402, among these also the APSE gene that encodes the CdtB-toxin.

In a second approach, we used weighted gene correlation network analysis (WGCNA) to identify modules of aphid or *H. defensa* genes that correlated to the *H. defensa* to aphid read ratio – an approximation of *H. defensa* titre – of each replicate (Table 2). For the WGCNA-

pipeline, filtering of genes is not recommended except for removing low-expression genes; after filtering out genes with less than 1 read per million the WGCA-approach included 8’600 aphid genes and 1’477 *H. defensa* genes. The approach clustered aphid genes into 18 modules and *H. defensa* genes into 12 modules.

We identified two aphid modules that correlated significantly positively with *H. defensa* titre, Aphid-w9 (r(13)=0.69, p=0.005) and Aphid-w10 (r(13)=0.64, p<0.001) (Supplementary Table 9). While no GO-terms were enriched in Aphid-w9, Aphid-w10 was associated with the GO-term ‘actin nucleation’. The WGCNA-approach also identified two *H. defensa* modules that correlated significantly with titre: Hdef-w11 (r(13)=0.81, p<0.001), in which no GO-terms were enriched, and Hdef-w8 (r(13)=0.77, p<0.001), in which the GO-term ‘type II secretion system (T2SS) complex’ was enriched. Targeted inspection of the expression of the T2SS genes showed, however, that this result was based on two T2SS-genes, *gspE* and *gspF*. Other T2SS genes, such as *gspD, gspL* and *gspM* were assigned to modules that did not correlate with titre. During the investigation we found that several genes of the T2SS, that were previously found in *H. defensa* of pea aphids (Chevignon et al. 2018), were not assembled from our sequencing data. Notably, H76 only expressed one out of five T2SS genes, *gspD*.

Correlation of primary and secondary endosymbiont gene expression

The same two correlation approaches as described above were applied to *Buchnera aphidicola* and *H. defensa* genes (Table 2). The strongest correlations were found between the *B. aphidicola* module BapHdef-B4 (no enriched GO-terms or KEGG pathways) and the two *H. defensa* modules BapHdef-H5 (r(13)=0.85, p<0.001), which contained the APSE gene encoding the CdtB-toxin and in which GO-terms such as ‘viral life cycle’ and ‘interaction with host’ were enriched, and BapHdef-H9 (r(13)=-0.85, p<0.001), in which the GO-term ‘macromolecule transmembrane transporter activity’ was enriched (Supplementary Figure 4, Supplementary Table 11).

The WGCNA approach identified one module of *B. aphidicola* genes, Bap-w6, whose eigengene’s expression correlated negatively (r(13)=-0.78, p=0.001) with *H. defensa* titre (Supplementary Table 11). No KEGG pathways or GO-terms were enriched in Bap-w6.

Characterisation of *Hamiltonella defensa* strains

To place our *H. defensa* strains in a phylogeny with other sequenced strains, 161 BUSCO genes were extracted from our transcriptome data and from publicly available and *H. defensa* genomes. Strain MED from *Bemisia tabaci* was used as an outgroup during phylogeny construction (Figure 7). Strains H15 and H85 were closely related and formed a separate clade that was well supported and basal to the other aphid-infecting strains we included. Strain H76 clustered with *H. defensa* A2C and AS3 from *A. pisum*, while strain H402 clustered with NY26 and 5AT from *A. pisum*.

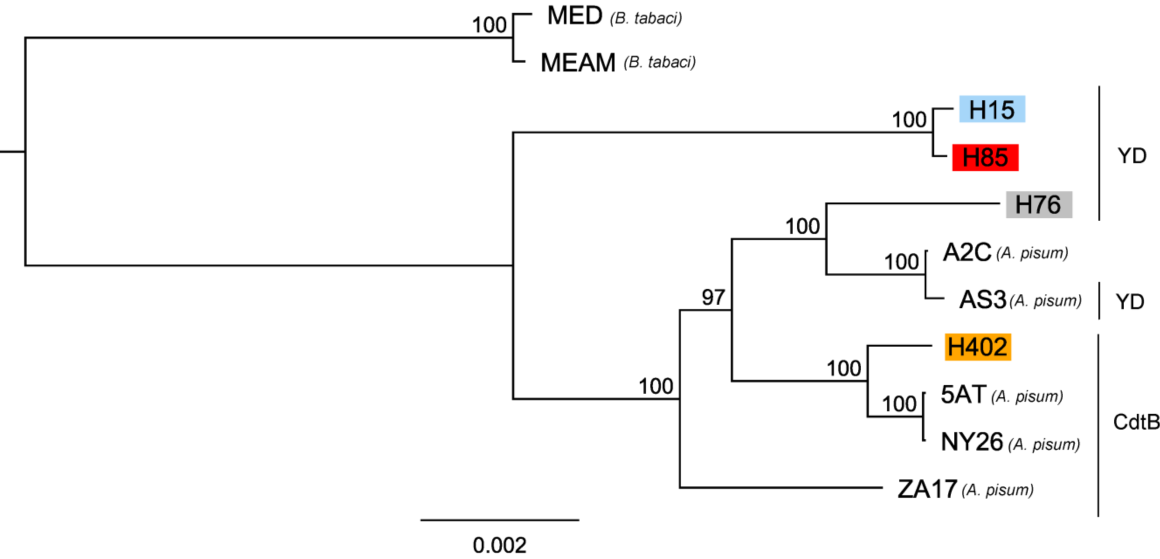


Figure 7 – *H. defensa* from different aphid species but with a similar APSE toxin cluster together. Phylogram based on 161 shared and complete BUSCO genes extracted from *H. defensa* strains of *A. fabae* (H15, H76, H85, H402), *Bemisia tabaci* (MED, MEAM) and *Acyrthosiphon pisum* (A2C, AS3, NY26, ZA17). Nodes are labelled with branch support based on percentage of bootstrap replicates that recovered the same node, and the toxin that the APSE encodes is indicated on the right (A2C has no APSE). Branch length is proportional to scale bar (unit: amino acid substitutions per site).

The APSE toxin cassettes of strains H76, H85 and H402 had already been sequenced (Dennis et al. 2017; Kerry Oliver, personal communication). The toxins assembled from the RNA-Seq data in this experiment confirmed our expectations from prior sequencing: Strain H85 carried a YD-repeat toxin that was identical to the reference toxin from H85. H15 carried the same toxin as H85. The YD-repeat toxin of H76 agreed with our expectations from the reference gene (NCBI GenBank: KU175898.1) but was longer and completed by a stop-codon. The CdtB toxin of H402 was retrieved from our data with one missense substitution (Glycine→Valine) compared to the reference gene (NCBI GenBank: KU175897.1). CdtB was also assembled in H15, H76 and H85 at low coverage, most likely due to index hopping (see Methods), and so was a partial YD-repeat toxin in H402.

Discussion

We used a triple RNA-Seq approach to monitor gene expression of the host *A. fabae* and its primary endosymbiont *B. aphidicola* in presence or absence of the secondary endosymbiont *H. defensa*. The four *H. defensa* strains used in the experiment are known to vary in the benefit and costs they inflict on the aphid host (Cayetano et al. 2015 and Supplementary Figure 1), show large variation in their gene expression and there is indication that the aphid host fails to control the density of one strain.

Host and *H. defensa* influence each other

The triple RNA-Seq approach was a necessity, as neither *B. aphidicola* nor *H. defensa* of *A. fabae* can currently be cultured, and advantageous as it allows investigating the interaction between the three different species through simultaneous analysis of gene expression in host and endosymbionts. Aphids and *H. defensa* seem to interact strongly with each other: The majority of *H. defensa* and aphid gene expression modules showed significant correlations with modules of the other species, and in four instances absolute correlation values were 0.9 or higher. Of immediate interest in further research will be the *H. defensa* gene module that showed a strong negative correlation with the hemocytin-containing aphid module. The genes contained in this *H. defensa* module might suppress the host's hemocytin-expression and thus allow the strain H85 to over-replicate. A candidate for suppression of hemocytin-expression is the *H. defensa* gene AS3p2_hypothetical_protein_CDS_BJP42_RS11500, which is encoded on a plasmid in *H. defensa* strain AS3 of pea aphids. The gene is strongly overexpressed in H85 only, and it is the DEG with the highest log2 fold change between the closely related strains H15 and H85.

H. defensa titre does not fully explain the impact on host gene expression

The high titre of H85 has curiously little impact on aphid gene expression: The host's gene expression reacted more strongly to presence of the intermediate-density H402 than H85. Additionally, only two aphid gene modules correlated with *H. defensa* titre, indicating that titre alone can only explain few expression changes in groups of co-expressed aphid genes.

Strain H85 impacts host hemocytes

Our results indicate that *H. defensa* H85 might impact the aphid's hemocytes. Speculatively, this could contribute to the severe costs of H85 on its host's survival and reproduction (Cayetano et al. 2015) if *H. defensa* H85 manages to escape the host's control by affecting hemocytes. However, how much hemocytes contribute to control of secondary endosymbionts

in aphids is not yet clear. Investigating *H. defensa* in pea aphids, Schmitz et al. (2012) found that hemocytes of pea aphids can take up bacterial endosymbionts but they could not conclude whether *H. defensa* cells were destroyed after phagocytosis or persisted in the lysosomes of the hemocytes. Our experiment provides a further puzzle piece suggesting that hemocytes might indeed be involved in controlling *H. defensa* density: the downregulation of hemocytin.

Hemocytin was first described in *Bombyx* (Kotani et al. 1995) and its homolog in *Drosophila* is called hemolectin (*hml*). Genes of the *hml* family include domains that are typically observed in vertebrate and arthropod clotting factors. They are important for coagulation after wounding (Goto et al. 2003; Scherfer et al. 2004) and in adult *D. melanogaster,* infection with Gram-negative bacteria leads to increased *hml* expression (Sanchez Bosch et al. 2019). However, our *A. fabae* clone did not significantly upregulate hemocytin expression in response to infection of with *H. defensa* strains H15, H76 or H402, and when infected with strain H85, the clone actually experienced strong downregulation of hemocytin (Figure 4 A). At first, we assumed that H85 might deplete the host's hemocytes. An endosymbiont-induced loss of hemocytes has been proposed in other insect-endosymbiont-systems; Garcia-Arraez et al. (2019) linked infection with the secondary endosymbiont *Spiroplasma poulsonii* to a lower number of sessile hemocytes and thus a reduced whole-body expression of *hml* in *D. melanogaster*, while Schmitz et al. (2012) found that *H. defensa* reduced numbers of sessile plasmatocytes in pea aphids. However, the expression of additional hemocyte markers indicated a change in the composition of the hemocyte population rather than hemocyte depletion. In *Drosophila,* the large majority of hemocytes express the two hemocyte-specific markers peroxidasin (*pxn*) and *hml*, while a minority only express either *pxn* or *hml* alone (Jung et al. 2005). If large-scale hemocyte depletion was responsible for the observed downregulation of hemocytin in aphids infected by H85, we would also expect *pxn* or other markers such as croquemort (*crq*), singed (*sn*), and lozenge (*lz*) to be downregulated. Since the expression of these markers was not significantly different in aphids infected with H85 compared to uninfected aphids or aphids infected with other *H. defensa* strains, the overall number of hemocytes likely did not change drastically. Instead, H85 might either have specifically inhibited expression of hemocytin or induced a shift from hemocytes expressing both *pxn* and hemocytin, to *pxn*-positive hemocytin-negative hemocytes. Shin et al. (2020) showed that in *Drosophila*, a decrease of the proportion of *pxn*-positive *hml*-negative hemocytes upon immunological or metabolic challenge is possible.

Even though the immune system of aphids generally responds ineffectively to Gram-negative bacteria (Gerardo et al. 2010; Guo et al. 2017; Laughton et al. 2011), co-occurrence of hemocytin downregulation and over-replication of H85 could suggest that the aphid's remaining immune response capacity might still be vital for controlling the density of the aphid's Gram-negative endosymbiont *H. defensa.* Our results motivate further studies analysing the reaction of the hemocytes to different *H. defensa* strains. Additionally, hemocytin expression and hemocyte numbers should be studied throughout the aphids life, since Laughton et al. (2014) showed that the impact of secondary endosymbionts on hemocyte numbers can change as the host ages.

Strain H402 activates a cluster of unknown genes

Another factor determining the host's reaction to a *H. defensa* strain may be the APSE it contains: In general, the YD-repeat-toxin encoding strains H15, H76 and H85 elicited less pronounced differential expression in the aphid host than the CdtB-toxin encoding H402. Presence of H402 was connected to strong differential expression (mostly upregulation) of 25 aphid genes with unknown function. Even though the aphid gene module containing these genes was linked to three *H. defensa* gene modules, in which GO-terms related to membrane composition and ATP synthesis were enriched, the mechanism behind the impact of strain H402 on these aphid genes of unknown function remains unclear. Nevertheless, our experiment points towards a clear difference in the aphid's reaction to the CdtB-toxin encoding strain H402 compared to all other strains.

B. aphidicola shows little reaction to presence of *H. defensa*

Even though *H. defensa* is known to require essential amino acids generated by the aphid's obligate endosymbiont *B. aphidicola* for survival (Degnan et al. 2009), *B. aphidicola* showed little reaction to the presence of *H. defensa*, both in terms of differential gene expression and density. The few *B. aphidicola* genes that were significantly differently expressed had low fold changes, and correlation between *B. aphidicola* and *H. defensa* gene modules was rarer and less strong than between aphid and *H. defensa* gene modules. These results are in agreement with previous studies and may not only reflect the sheltered intracellular lifestyle of *B. aphidicola* (Ratzka et al. 2012) but also the reduced potential of *B. aphidicola* to alter its gene expression due to the loss of regulatory genes (Moran et al. 2005b; Wilcox et al. 2003).

APSE activity can explain some but not all differences between *H. defensa* strains

Based on the analysis of BUSCO genes, the two *H. defensa* strains H76 and H402 were closely related to strains infecting pea aphids with same-type APSE-toxins, while strains H15 and H85 were different from any of the sequenced strains from *A. pisum*. Our phylogeny is in agreement with the phylogenies for *H. defensa* and APSE in Rouïl et al. (2020), even though we used considerably more genes. Given the short evolutionary distance between the strains H15 and H85, their very different phenotypes – in terms and costs and density that they induce or reach in the host – are even more intriguing.

The three *H. defensa* haplotypes – haplotype 1 comprising of strain H76, haplotype 2 of H402 and haplotype 3 of H15 and H85 (Cayetano et al. 2015) – displayed markedly different gene expression patterns. Compared to H85, strains H76 and H402 differentially expressed genes enriched for GO-terms 'pathogenesis' and 'interspecific interaction'. Notably APSE genes were more active in H76 and H402 than in H85 and H15. Since APSE is lysogenic (Degnan and Moran 2008; Moran et al. 2005a; van der Wilk et al. 1999) one could assume that higher APSE activity reduces a strains' density and thus the cost in terms of resources that the strains divert from the host. Therefore the higher APSE activity in H76 and H402 might explain why these strains are less costly than H85. Yet, APSE activity cannot explain the different densities of H15 and H85, as no APSE-genes were differentially expressed between the two strains. Instead, GO-terms linked to ribosomes were enriched in the DEG between the two *H. defensa* H15 and H85. Differential regulation of ribosomal proteins is not well understood, but has been found to be associated with stress or different growth conditions in bacteria and yeast (Gasch et al. 2000; Tao et al. 1999; Zhou et al. 2010) and other taxa (Zhou et al. 2015). Given the over-replication of H85, stress could be induced either by limited nutrient availability or interaction with the aphid's immune system.

Conclusions

1) Variation in gene expression indicates differences in the mechanism underlying the cost to the host induced by different strains of *H. defensa*.

2) The over-replicating strain H85 impacts the aphid's hemocytin expression, suggesting experimental investigation of the link between hemocytin and *H. defensa* in aphids.

3) While there are strong correlations between aphid and *H. defensa* gene modules, which implies an interspecific interaction, presence of *H. defensa* impacts gene expression of the aphid's obligate endosymbiont (*Buchnera aphidicola*) much less.

Methods

Aphid clones

This study uses a subset of the 12 sublines of the *A. fabae* clone A06-407 described by Cayetano et al. (2015). Clone A06-407 was naturally free of secondary endosymbionts and was infected with *H. defensa* by microinjection of hemolymph from other *A. fabae* clones between 2008 and 2012. In this experiment, we used four of these *H. defensa*-infected sublines: H15, H76, H85 and H402. Collection details of the clone A06-407 and the *H. defensa*-infected hemolymph donors are provided in Supplementary Table 12. Sublines were maintained on broad beans (*Vicia faba*) under environmental conditions ensuring clonal reproduction (16 h photoperiod at 18-20°C). For each subline, 12 bean seedlings were infested with adult aphids. After reproduction, DNA of the adults was extracted to confirm aphid clone identity, presence and haplotype of *H. defensa*. To avoid environmental maternal effects carrying over from stock cultures, the 12 colonies per subline were maintained at 18°C for two generations. Nymphs of the third generation were reared at 22°C for eight days until adult. The 12 plants per subline were divided into two batches (A and B) of six plants each and replicates of 18 aphids (R1 and R2 from plant batch A, R3 and R4 from plant batch B) were collected, with each plant of a batch contributing three aphids.

Secondary endosymbiont strains

The four *H. defensa* strains – H15, H76, H85 and H402 – comprise three haplotype groups. For some of the strains, the APSE toxin cassettes were previously sequenced, showing that different APSE variants are present in *A. fabae* (Dennis et al. 2017). H76 belongs to the *H. defensa* haplotype 1. It carries an APSE that encodes a YD-repeat toxin gene with two open reading frames (NCBI GenBank: KU175898). The protection against the parasitoid *Lysiphlebus fabarum* provided by H76 has been found by Cayetano et al. (2015) to be very strong, while aphids infected by H76 were virtually as fecund as uninfected controls (Supplementary Figure 1). Strain H402 belongs to haplotype 2. It carries an APSE that encodes a CdtB-toxin (NCBI GenBank: KU175897). The protection provided by H402 to A06-407 is intermediate, and so are its costs (Cayetano et al. 2015). Strains H15 and H85 belong to haplotype 3. H85 carries an APSE encoding a YD-repeat toxin gene that is longer

than the one of H76. For H15 the APSE toxin was not sequenced prior to this experiment. Both H15 and H85 provide limited protection and entail high costs (Cayetano et al. 2015). Strain H85 is particularly costly: Aphids infected with H85 die shortly after reaching adulthood. In contrast to H15, H85 reaches very high density in the host (Cayetano et al. 2015).

RNA extraction

Aphids were crushed in 0.5 ml TRIzol (Thermo Fisher). The volume of TRIzol was adjusted to 1 ml and after 20 min at room temperature (RT), the samples were stored at -80°C. All further procedures up to library quantification were conducted successively on the two batches A, comprising replicates R1 and R2, and B, comprising replicates R3 and R4.

Samples were thawed at RT and vigorously shaken by hand with 200 µl chloroform (PanReac AppliChem). After incubation for 10 min at RT and centrifugation at 4°C, the aqueous supernatant was recovered and re-extracted with chloroform. RNA was pelleted by centrifugation after mixing with 500 µl ice-cold isopropanol (Merck) and incubating for 4 hrs at -20°C. The pellet was air-dried at RT, washed with ice-cold 75% ethanol and absolute ethanol and re-suspended in 50 µl RNase-free water.

DNA was removed using the RNase-free DNase kit (Qiagen) and RNA was purified using the RNeasy Mini Kit (Qiagen). Cleaned RNA was eluted in 30 µl RNase-free water and RNA integrity was assessed with the RNA 6000 Nano kit (Agilent) on the Bioanalyzer 2100 (Agilent) (Supplementary Figure 5). To deplete ribosomal RNA (rRNA) but maintain bacterial mRNA, 1 µg total RNA was processed with the riboZero Epidemiology kit (Illumina), using half of the recommended volume of reagents per reaction. Sample volume was adjusted to 180 µl and RNA was recovered with glycogen-assisted ethanol precipitation.

Library preparation

RNA pellets were re-suspended in 9 µl Fragment, Prime, Finish mix from the TruSeq stranded mRNA library preparation kit (Illumina). RNA was fragmented at 94°C for 95 seconds and 8.5 µl were processed according to the manual of the TruSeq kit, using half of the recommended reagent amounts and starting at the “Synthesize First Strand cDNA” step. Eleven PCR cycles were enough to achieve the desired library amplification. For barcoding, all 12 barcodes from the TruSeq RNA Single Indexes Set A (Illumina) and eight additional barcodes from Set B were used. Libraries were size selected with Agencourt AMPure XP

beads (Beckman Coulter). The bead pellet was washed twice with 200 µl 80% ethanol, dried at RT and eluted in 15 µl Resuspension Buffer from the TruSeq kit. Average library fragment length, determined with the High Sensitivity DNA Analysis Kit (Agilent), varied between 421-460 bp, with exception of H15R1 (385 bp). Library concentrations were quantified on a Roche LightCycler 480 using the Universal Kapa library quantification kit (Kapa Biosystems).

Transcriptome sequencing

A pool containing 10 nM of each library was sequenced by the Functional Genomics Center Zurich on an Illumina HiSeq 4000 (one lane, 150 bp paired-end reads). The number of usable reads per library surviving preliminary trimming and rRNA removal was recorded. For a second sequencing run, we designed a new library pool to approximately balance the number of usable reads per library.

De novo Assembly

To prepare reads for *de novo* assembly, sequencing adapters, potential primer mismatches and stretches of low-quality bases were removed from the reads using Trimmomatic v0.35 (Bolger et al. 2014) with default settings except for a sliding window with a minimal average quality of 20, a minimal read length of 100 bp and a 6 bp headcrop. To prepare the reads for mapping, the minimal average quality trimming threshold was relaxed to 15, headcrop was deactivated, and the minimal read length was set to 75 bases. Only reads surviving in pairs were used for subsequent steps. Kraken v0.10.5 (Wood and Salzberg 2014) indicated significant human RNA contamination in library H15R1 and minor contamination in libraries H15R3 and H15R4. Reads prepared for *de novo* assembly and mapping were quality checked using FastQC v0.11.4 (Andrews 2010).

To recover *B. aphidicola* and aphid transcripts, the 17 uncontaminated libraries were assembled in Trinity v2.1.1 (Grabherr et al. 2011) with *in silico* read normalisation to a maximal coverage of 50 and requiring a minimal transcript length of 200. The assembled transcripts were clustered with CD-HIT-EST from the CD-HIT suite v4.6.5 (Fu et al. 2012) with a sequence identity threshold of 0.95, a band width of 50, and clustering to the most similar cluster (g=1). Transcripts that did not reach a normalised expression metric of 0.5 TPM (transcripts per million transcripts) in at least one library were removed with the Trinity v2.4.0 script filter_low_expr-transcripts.pl. Ribosomal sequences were removed with Ribopicker v0.4.3 (Schmieder et al. 2012) and polyA-tails longer than 5 bases were trimmed

with prinseq v0.20.14 (Schmieder and Edwards 2011). The transcripts were assigned to the most likely organism of origin using blastn v2.2.30 (Altschul et al. 1997; Camacho et al. 2009) against custom databases with default settings except for an E-value cutoff of $1e^{-8}$ and a maximal number of HSPs (alignments) of 1. Transcripts that could not be assigned to an organism of origin were blasted against the nr database using diamond v0.9.22 (Buchfink et al. 2014) and an E-value cutoff of $1e^{-5}$. Transcripts that blasted to more than one organism were discarded if the bitscore difference between best hits to different taxa was less than 100 or assigned to the best-scoring taxa if the bitscore difference was more than 100 using a custom script. Transcripts assigned to either aphids or *B. aphidicola* were separated and clustered to a sequence identity of 0.9 using CD-HIT-EST as described above.

To recover transcripts of *H. defensa* and its associated APSE, we performed *de novo* assembly for each *H. defensa* strain separately. For each assembly, four libraries containing the respective strain were combined, except for H15 where we excluded the heavily contaminated library H15R1. For each strain-specific assembly we retained transcripts blasting to *H. defensa* and its associated virus APSE using the same procedures and cutoffs as described above.

De novo-assembled bacterial transcripts represent operons and likely contain multiple genes. Since genes that lie on the same operon can be differentially expressed (Conway et al. 2014), the signal of a differentially expressed gene (DEG) may be diluted in transcript level analyses if other genes on the same operon are not differentially expressed. Additionally, inefficient transcription termination between convergent operons (Conway et al. 2014) could *in silico* merge functionally unrelated operons. To avoid such artefacts, differential gene expression in bacteria had to be assessed at gene instead of transcript level.

To identify *B. aphidicola* genes, annotations from the closely related reference genome of *B. aphidicola* from *Aphis glycines* (NCBI GenBank: NZ_CP009253.1, NZ_CP009254.1, NZ_CP009255.1) were transferred to the *de novo* assembled transcripts of *B. aphidicola*. This was achieved by aligning transcripts to the reference genome in Geneious v11.0.5 (Kearse et al. 2012). Manual correction steps included removal of ten chimeric *B. aphidicola* transcripts and inclusion of one unaligned transcript that matched the genome of *B. aphidicola* of *Uroleucon ambrosiae* (NCBI GenBank: CP002648.1). Sequence differences between reference genome and aligned transcripts were corrected manually and gene start and stop

sites were adjusted where supported by reads. After removal of rRNA genes, annotated genes were exported for downstream analysis.

Identifying *H. defensa* genes needed a different approach, as only half of *H. defensa* and APSE transcripts aligned to related genomes. Instead of attempting the transfer of annotations, we predicted the genes from the strain-specific *H. defensa* transcriptomes: We ran Prokka v1.11 (Seemann 2014) while providing it with known *H. defensa* proteins (Supplementary Table 13). For each strain the predicted genes were clustered to a sequence identity of 0.8 using a length difference cutoff of 0.9 and deduplicated using a sequence identity of 0.99 and a length difference cutoff of 0.1 with CD-HIT-EST. The four gene sets were pooled, and annotations were manually curated so that strain-specific variants of the same genes were labelled identically. Gene variants were aligned with Geneious to detect and remove chimeras.

In RNA-Seq studies comparing the expression of several related bacterial strains, the correct choice of reference genome is crucial as phylogenetic distance can lead to false-positives in the differential expression analysis (Price and Gibas 2017). Since Kallisto cannot interpret non-ATGCU bases, use of a consensus sequence from all four strains would have resulted in the replacement of 29’074 or 2.36% of all *H. defensa* bases with pseudo-random bases. To avoid such a high percentage of pseudo-random bases, we used the consensus sequence of strains H15 and H85 for differential expression analysis in *H. defensa*. This decreased the number of pseudo-random bases inserted by Kallisto to 514. The differential expression results for *H. defensa* are therefore most accurate for the two strains H15 and H85, while there may be some false positives due to phylogenetic distance for strains H76 and H402.

Differential expression analysis

Using Kallisto v0.43.0 (Bray et al. 2016), reads trimmed for mapping were aligned simultaneously to annotated genes from *B. aphidicola*, consensus sequences of *H. defensa* and APSE genes, transcripts of *A. fabae,* and the coding DNA sequences (CDS) of the most frequent contaminant bacteria (Supplementary Table 14). The resulting abundance tables were split with a custom R function to allow organism-specific read normalisation during differential gene expression analysis using the package DESeq 2 v1.22.1 (Love et al. 2014) and tximport v1.10.0 (Soneson et al. 2016), which provided the read counts to DESeq 2, in R v3.3.2 (R Core Team 2018). For the analysis, aphid transcripts were merged to gene level. From the assembly 46’352 transcripts resulted, of which 2’590 did not blast to any known

record in NCBI and were discarded from differential expression analysis as they might correspond to chimeric assembly artefacts or non-coding RNA. The remaining transcripts corresponded to 19'864 different Trinity 'genes' that were annotated as 10'809 different genes. It is important to note that Trinity assigns the 'gene' status purely based on mathematical, not biological, information. For the differential expression analysis, we therefore merged all Trinity genes that were annotated as the same gene. For this, the list of annotations was manually curated, removing differences in gene annotations such as the terms 'Predicted' or 'isoform X1'.

Differential expression analysis was done separately for each species. For analysing aphid gene expression, genes with less than 1 read per million were discarded from analysis. To compare differences in expression of aphid genes between two treatments in DESeq 2, we used Wald tests. The differential expression models contained the two variables *batch* (A, B) and *treatment* (H0, H15, H76, H85, H402) as fixed factors. Based on AICc values, the model without interaction of the fixed factors was used. We provided the DESeq 2 result function with two significance thresholds: adjusted p-value alpha <0.01 and log fold change >0.25. We drew pairwise comparisons between aphid gene expression in the presence of a *H. defensa* strain (H15, H76, H85 or H402) to gene expression in absence of *H. defensa* (H0). Additionally, we compared aphid gene expression in presence of strain H15 to gene expression in presence of H85 using a reduced model, in which we only included libraries of treatments H15 and H85. Differential expression analysis as described above was repeated for *B. aphidicola* genes. For *H. defensa* genes, gene expression of strains H15, H76 and H402 was compared to gene expression of strain H85. A high number of all *H. defensa* genes was differentially expressed between both H76 and H402 and the reference strain H85. Since a basic assumption of differential expression analysis is that most genes are not differentially expressed (Dillies et al. 2013), this could have corrupted the differential expression model. Thus, we used a reduced model containing only libraries of treatments H15 and H85 to confirm the DEG between strains H15 and H85.

We used UpSetR v1.3.3. (Lex et al. 2014) to visualise the number of DEG shared between treatments and pcaExplorer v2.8.0 (Marini and Binder 2019) to visualise the results of principal component analyses (PCA), which used the expression patterns of all genes of each organism to segregate the libraries, showing the ordination along the two first axes.

Aphid and endosymbiont genes were annotated with GO-terms using Blast2GO v5.2.5 (Götz et al. 2008). The DEG of each treatment were tested for GO-term enrichment using two-tailed Fisher exact test while filtering for FDR≤0.05 using OmicsBox (BioBam Bioinformatics 2019). To prevent genes with many isoforms biasing this analysis, we randomly selected one isoform per gene for GO-term enrichment.

Phylogenetic tree

For phylogenetic analysis of *H. defensa,* we predicted BUSCO genes from the strain-specific gene sets as well as from the CDS from known and reasonably complete *H. defensa* genomes (accession numbers in Supplementary Table 13) with BUSCO v3.0.2 (Kriventseva et al. 2017; Kriventseva et al. 2015) using the proteobacteria dataset as reference. Single-copy complete BUSCO genes present in all strains were extracted, translated to protein sequences and aligned with MAFFT v7.273 (Katoh and Standley 2013). The per-gene alignments were trimmed using trimAl v1.2 rev59 (Silla-Martínez et al. 2009) and concatenated into a superalignment with Geneious. Genes with obvious frameshifts or truncations were removed, reducing the number of shared BUSCO genes from 164 to 161. Best-fit partitioning schemes and models of evolution were selected with PartitionFinder v2.1.1 (Lanfear et al. 2016) using RAxML (Stamatakis 2014) and the relaxed clustering algorithm (Lanfear et al. 2014). Each gene corresponded to one data block, and models of evolution were selected based on AICc. Note that we did not consider models of evolution with equal base frequencies, or base frequencies determined using maximum likelihood, or amino acid frequencies estimated from mitochondrial, chloroplast, HIV viral or influenza viral datasets. The phylogenetic tree was generated with RAxML v8.2.11 in Geneious executing 500 rapid bootstrap interferences followed by a Maximum Likelihood search on data partitioned according to the PartitionFinder results.

Correlation of host and symbiont expression

Aphid and *B. aphidicola* gene expression were correlated to *H. defensa* gene expression using two approaches, the weighted correlation network analysis approach (WGCNA) (Langfelder and Horvath 2008; Zhang and Horvath 2005) and the correlation approach described in Smith and Moran (2020). For the WGCNA approach, genes with a read count less than 1 read per million were discarded and read counts were log-transformed using the rld function. We followed the procedures described in Smith and Moran (2020) except for using signed hybrid networks and the biweight midcorrelation as a robust alternative to Pearson correlation.

Briefly, modules of co-expressed genes were constructed from the rld-transformed expression data using hierarchical clustering. Module eigengenes (defined as the first principal component of a module) were calculated and correlated with *H. defensa* titre (approximated by the *H. defensa* to aphid read ratio).

For the correlation approach, invariant genes with interquartile ranges (IQR) ≤0.15 for aphid and ≤0.75 for *H. defensa* were removed based on inspection of histograms. A gene correlation matrix with the Pearson's correlation coefficient of each pairwise combination of IQR-filtered aphid and *H. defensa* genes was constructed. Genes were clustered by their expression patterns into modules using flashClust v.1.01-2 (Langfelder and Horvath 2012) and similar modules were combined after estimating an adequate number of clusters using the "gap" statistic implemented in cluster v.2.0.7-1 (Maechler et al. 2018) and inspecting module correlation dendrograms. To test whether the pattern of gene clustering was due to random chance, the simprof similarity profile permutation test implemented in clustsig v 1.1 (Whitaker and Christman 2014) was used. It created an expected data distribution from 100 similarity profiles and compared the observed test statistics to the null distribution based on 99 similarity profiles at $\alpha<0.01$. Module eigengenes were calculated with WGCNA v. 1.66 (Langfelder and Horvath 2008; Zhang and Horvath 2005) and were correlated in an eigengene correlation matrix as well as to titre (approximated by the *H. defensa* to aphid read ratio).

Both WGCNA and correlation gene modules were tested for enrichment of GO-terms using GoFuncR v1.6.1 (Grote 2020) in R v3.6.3 and results were corrected for multiple testing and interdependency with 1000 replicates and an adjusted significance threshold of $q<0.01$. *B. aphidicola* modules were additionally tested for enrichment of KEGG-pathways with clusterprofiler v3.14.3 (Yu et al. 2012) in R v3.6.3.

Index hopping

Even though aphids in the control treatment (H0) were free from secondary endosymbionts, we registered expression of *H. defensa* genes in H0. The observation could be explained by contamination, misassignment of reads between genes that are conserved between *B. aphidicola* and *H. defensa*, or index hopping, with the last being the most parsimonious explanation. Firstly, misassignment to conserved genes was not the cause: Highly expressed and strain-specific genes like the CdtB and YD-repeat toxin seemed to be expressed at low levels in *H. defensa* strains that did not contain these genes as well as in the 'phantom' *H. defensa* of H0 aphids. Secondly, index hopping was more likely than contamination for

two reasons. A) Index hopping is expected to occur in sequencing assays like the one we used (Illumina Inc 2018). B) Contamination of every single library with foreign *H. defensa* is unlikely. We therefore accepted index hopping as the most parsimonious explanation.

Declarations

Author's contributions

HK, ABD and CV designed research, HK performed research and analysed data, and HK and CV wrote the paper.

Acknowledgements

We thank the team of the Functional Genomics Center Zurich (FGCZ) for sequencing, the Genetic Diversity Centre (GDC) at ETH Zurich for support during data analysis, and Paula Rodriguez for the long-term maintenance of the insect lines.

Availability of data and materials

Raw transcriptomic data has been deposited in GenBank, under BioSample ID no. SAMN10606880 to AMN10606919.

Competing interests

The authors declare no competing interests.

Funding

This project has been funded by the Swiss National Science Foundation (Sinergia Grant No. CRSII3_154396).

Literature cited

Altincicek B, ter Braak B, Laughton AM, Udekwu KI, Gerardo NM. 2011. *Escherichia coli* K-12 pathogenicity in the pea aphid, *Acyrthosiphon pisum*, reveals reduced antibacterial defense in aphids. Developmental & Comparative Immunology. 35(10):1091-1097.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25(17):3389-3402.

Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Barribeau SM, Parker BJ, Gerardo NM. 2014. Exposure to natural pathogens reveals costly aphid response to fungi but not bacteria. Ecology and Evolution. 4(4):488-493.

BioBam Bioinformatics. 2019. OmicsBox – Bioinformatics Made Easy. https://www.biobam.com/omicsbox.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114-2120.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 34(5):525-527.

Brinza L, Viñuelas J, Cottret L, Calevro F, Rahbé Y, Febvay G, Duport G, Colella S, Rabatel A, Gautier C et al. 2009. Systemic analysis of the symbiotic function of *Buchnera aphidicola*, the primary endosymbiont of the pea aphid *Acyrthosiphon pisum*. Comptes Rendus Biologies. 332(11):1034-1049.

Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. Nature Methods. 12:59.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. Bmc Bioinformatics. 10:421-421.

Cayetano L, Rothacher L, Simon J-C, Vorburger C. 2015. Cheaper is not always worse: strongly protective isolates of a defensive symbiont are less costly to the aphid host. Proceedings of the Royal Society B. 282(1799).

Chevignon G, Boyd BM, Brandt JW, Oliver KM, Strand MR. 2018. Culture-Facilitated Comparative Genomics of the Facultative Symbiont *Hamiltonella defensa*. Genome Biol Evol. 10(3):786-802.

Conway T, Creecy JP, Maddox SM, Grissom JE, Conkle TL, Shadid TM, Teramoto J, San Miguel P, Shimada T, Ishihama A et al. 2014. Unprecedented High-Resolution View of Bacterial Operon Architecture Revealed by RNA Sequencing. Mbio. 5(4):e01442-01414.

Costello M, Fleharty M, Abreu J, Farjoun Y, Ferriera S, Holmes L, Granger B, Green L, Howd T, Mason T et al. 2018. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. BMC genomics. 19(1):332-332.

Degnan PH, Moran NA. 2008. Diverse Phage-Encoded Toxins in a Protective Insect Endosymbiont. Appl Environ Microb. 74(21):6782-6791.

Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA. 2009. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. P Natl Acad Sci USA. 106(22):9063-9068.

Dennis AB, Patel V, Oliver KM, Vorburger C. 2017. Parasitoid gene expression changes after adaptation to symbiont-protected hosts. Evolution. 71(11):2599-2617.

Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J et al. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Briefings in Bioinformatics. 14(6):671-683.

Douglas AE. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. Annual Review of Entomology. 43(1):17-37.

Dykstra HR, Weldon SR, Martinez AJ, White JA, Hopper KR, Heimpel GE, Asplen MK, Oliver KM. 2014. Factors limiting the spread of the protective symbiont *Hamiltonella defensa* in *Aphis craccivora* aphids. Appl Environ Microb. 80(18):5818-5827.

Franc NC, Dimarcq J-L, Lagueux M, Hoffmann J, Ezekowitz RAB. 1996. *Croquemort*, A Novel Drosophila Hemocyte/Macrophage Receptor that Recognizes Apoptotic Cells. Immunity. 4(5):431-443.
Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 28(23):3150-3152.
Garcia-Arraez MG, Masson F, Escobar JCP, Lemaitre B. 2019. Functional analysis of RIP toxins from the *Drosophila* endosymbiont *Spiroplasma poulsonii*. BMC Microbiology. 19(1):46.
Garschall K, Flatt T. 2018. The interplay between immunity and aging in *Drosophila* [version 1; peer review: 2 approved]. F1000Research. 7(160).
Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell. 11(12):4241-4257.
Gehrer L, Vorburger C. 2012. Parasitoids as vectors of facultative bacterial endosymbionts in aphids. Biology letters. 8(4):613-615.
Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, Duncan EJ, Evans JD, Gabaldon T, Ghanim M et al. 2010. Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*. Genome Biol. 11(2):R21.
Goto A, Kadowaki T, Kitagawa Y. 2003. *Drosophila* hemolectin gene is expressed in embryonic and larval hemocytes and its knock down causes bleeding defects. Dev Biol. 264(2):582-591.
Goto A, Kumagai T, Kumagai C, Hirose J, Narita H, Mori H, Kadowaki T, Beck K, Kitagawa Y. 2001. A *Drosophila* haemocyte-specific protein, hemolectin, similar to human von Willebrand factor. Biochem J. 359(Pt 1):99-108.
Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research. 36(10):3420-3435.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology. 29(7):644-652.
Grote S. 2020. GOfuncR: Gene ontology enrichment using FUNC. R package version 1.8.0.
Guo J, Hatt S, He K, Chen J, Francis F, Wang Z. 2017. Nine facultative endosymbionts in aphids. A review. Journal of Asia-Pacific Entomology. 20(3):794-801.
Hafer-Hahmann N, Vorburger C. 2020. Parasitoids as drivers of symbiont diversity in an insect host. Ecology Letters. 23:1232-1241.
Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. P Natl Acad Sci USA. 108(7):2849-2854.
Effects of Index Misassignment on Multiplexing and Downstream Analysis. 2018. 770-2017-004-D. [accessed 2020 February]. https://emea.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf.
Jung S-H, Evans CJ, Uemura C, Banerjee U. 2005. The *Drosophila* lymph gland as a developmental model of hematopoiesis. Development. 132(11):2521.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution. 30(4):772-780.
Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12):1647-1649.
Koga R, Meng XY, Tsuchida T, Fukatsu T. 2012. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. P Natl Acad Sci USA. 109(20):E1230-E1237.
Kotani E, Yamakawa M, Iwamoto S, Tashiro M, Mori H, Sumida M, Matsubara F, Taniai K, Kadono-Okuda K, Kato Y. 1995. Cloning and expression of the gene hemocytin, an insect humoral lectin which is homologous with the mammalian von Willebrand factor. Biochimica et biophysica acta. 1260(3):245-258.

Kriventseva EV, Simão FA, Klioutchnikov G, Seppey M, Manni M, Ioannidis P, Waterhouse RM, Zdobnov EM. 2017. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Molecular Biology and Evolution. 35(3):543-548.

Kriventseva EV, Zdobnov EM, Simão FA, Ioannidis P, Waterhouse RM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31(19):3210-3212.

Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. 2014. Selecting Optimal Partitioning Schemes for Phylogenomic Datasets Bmc Evol Biol. 14(1).

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2016. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Molecular Biology and Evolution. 34(3).

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. Bmc Bioinformatics. 9(1):559.

Langfelder P, Horvath S. 2012. Fast R Functions for Robust Correlations and Hierarchical Clustering. Journal of Statistical Software. 46(11):1-17.

Laughton AM, Fan MH, Gerardo NM. 2014. The combined effects of bacterial symbionts and aging on life history traits in the pea aphid, *Acyrthosiphon pisum*. Appl Environ Microb. 80(2):470-477.

Laughton AM, Garcia JR, Altincicek B, Strand MR, Gerardo NM. 2011. Characterisation of immune responses in the pea aphid, *Acyrthosiphon pisum*. J Insect Physiol. 57(6):830-839.

Lebestky T, Chang T, Hartenstein V, Banerjee U. 2000. Specification of *Drosophila* Hematopoietic Lineage by Conserved Transcription Factors. Science. 288(5463):146.

Lemaitre B, Hoffmann J. 2007. The Host Defense of *Drosophila melanogaster*. Annual Review of Immunology. 25(1):697-743.

Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: Visualization of Intersecting Sets. IEEE Transactions on Visualization and Computer Graphics. 20(12):1983-1992.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12):550-550.

Maechler M, Rousseeuw p, Struyf A, Hubert M, Hornik K. 2018. cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1.

Marini F, Binder H. 2019. pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. Bmc Bioinformatics. 20(1):331.

Moran N, Baumann P. 1994. Phylogenetics of cytoplasmically inherited microorganisms of arthropods. Trends in Ecology & Evolution. 9(1):15-20.

Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H. 2005a. The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes. P Natl Acad Sci USA. 102(47):16919-16926.

Moran NA, Dunbar HE, Wilcox JL. 2005b. Regulation of transcription in a reduced bacterial genome: nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. J Bacteriol. 187(12):4229-4237.

Moran NA, Russell JA, Koga R, Fukatsu T. 2005c. Evolutionary Relationships of Three New Species of Enterobacteriaceae Living as Symbionts of Aphids and Other Insects. Appl Environ Microb. 71(6):3302-3310.

Nelson RE, Fessler LI, Takagi Y, Blumberg B, Keene DR, Olson PF, Parker CG, Fessler JH. 1994. Peroxidasin: a novel enzyme-matrix protein of *Drosophila* development. The EMBO Journal. 13(15):3438-3447.

Oliver KM, Campos J, Moran NA, Hunter MS. 2008. Population dynamics of defensive symbionts in aphids. 275(1632):293-299.

Oliver KM, Degnan PH, Burke GR, Moran NA. 2010. Facultative Symbionts in Aphids and the Horizontal Transfer of Ecologically Important Traits. Annual Review of Entomology. 55(1):247-266.

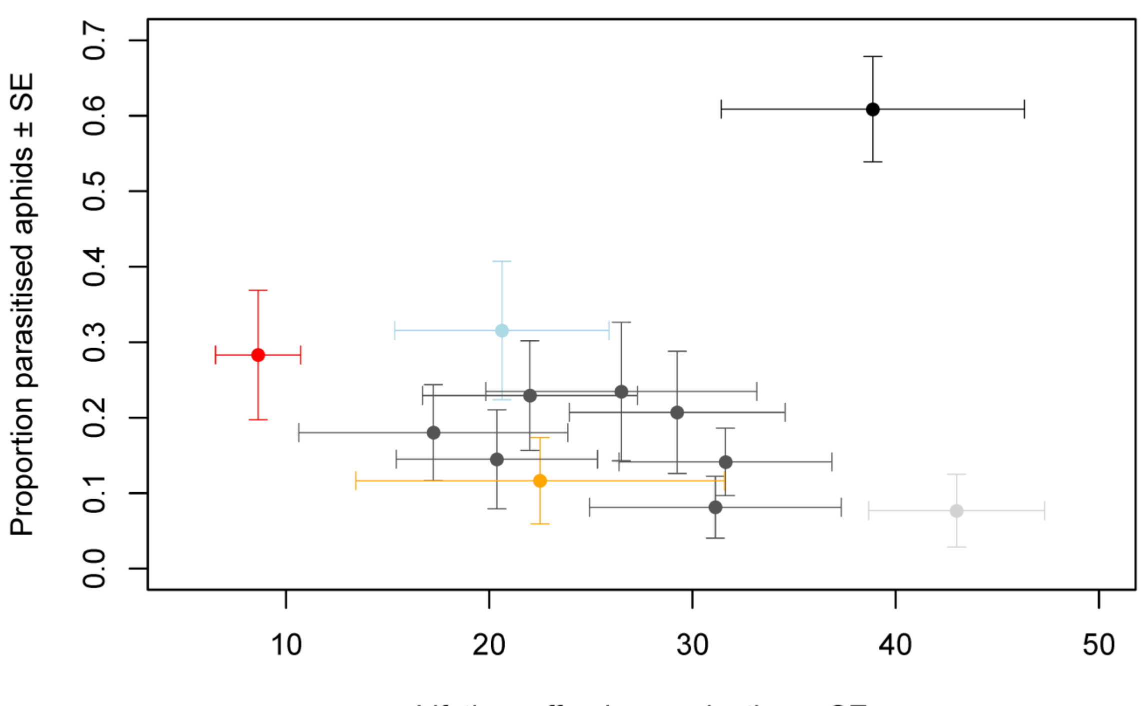
Oliver KM, Degnan PH, Hunter MS, Moran NA. 2009. Bacteriophages Encode Factors Required for Protection in a Symbiotic Mutualism. Science. 325(5943):992-994.

Oliver KM, Higashi CH. 2019. Variations on a protective theme: *Hamiltonella defensa* infections in aphids variably impact parasitoid success. Current opinion in insect science. 32:1-7.

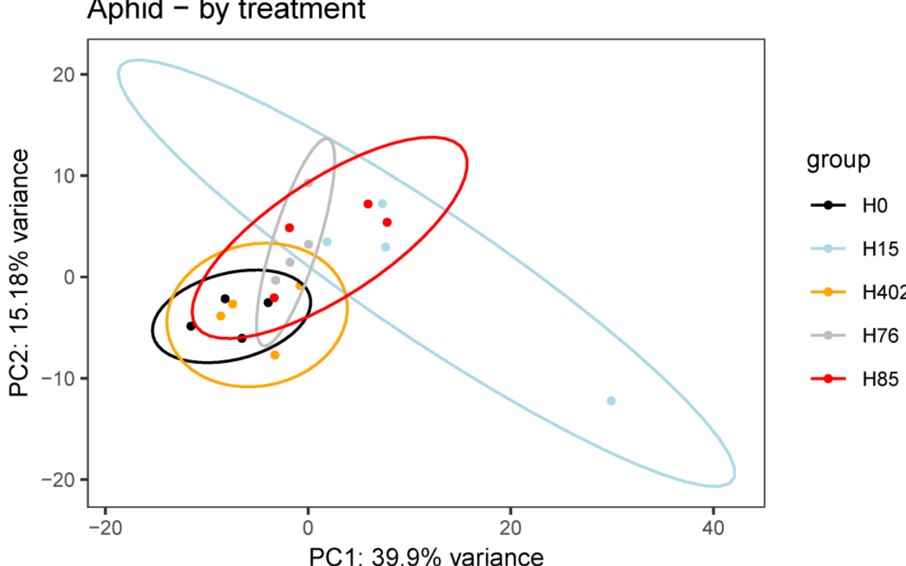
Oliver KM, Moran NA, Hunter MS. 2005. Variation in resistance to parasitism in aphids is due to symbionts not host genotype. P Natl Acad Sci USA. 102(36):12795-12800.
Oliver KM, Russell JA, Moran NA, Hunter MS. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proceedings of the National Academy of Sciences. 100(4):1803-1807.
Price A, Gibas C. 2017. The quantitative impact of read mapping to non-native reference genomes in comparative RNA-Seq studies. PLOS ONE. 12(7):e0180904.
R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Ratzka C, Gross R, Feldhaar H. 2012. Endosymbiont Tolerance and Control within Insect Hosts. Insects. 3(2):553-572.
Renoz F, Noel C, Errachid A, Foray V, Hance T. 2015. Infection Dynamic of Symbiotic Bacteria in the Pea Aphid *Acyrthosiphon pisum* Gut and Host Immune Response at the Early Steps in the Infection Process. PLoS One. 10(3):e0122099.
Rouïl J, Jousselin E, Coeur d'Acier A, Cruaud C, Manzano-Marín A. 2020. The protector within: Comparative genomics of APSE phages across aphids reveals rampant recombination and diverse toxin arsenals. Genome Biol Evol.
Russell JA, Moran NA. 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. Proceedings of the Royal Society B. 273(1586):603-610.
Sanchez-Contreras M, Vlisidou I. 2008. The Diversity of Insect-bacteria Interactions and its Applications for Disease Control. Biotechnology and Genetic Engineering Reviews. 25(1):203-244.
Sanchez Bosch P, Makhijani K, Herboso L, Gold KS, Baginsky R, Woodcock KJ, Alexander B, Kukar K, Corcoran S, Jacobs T et al. 2019. Adult *Drosophila* Lack Hematopoiesis but Rely on a Blood Cell Reservoir at the Respiratory Epithelia to Relay Infection Signals to Surrounding Tissues. Developmental Cell. 51(6):787-803.e785.
Scarborough CL, Ferrari J, Godfray HCJ. 2005. Aphid Protected from Pathogen by Endosymbiont. Science. 310(5755):1781.
Scherfer C, Karlsson C, Loseva O, Bidla G, Goto A, Havemann J, Dushay MS, Theopold U. 2004. Isolation and Characterization of Hemolymph Clotting Factors in *Drosophila melanogaster* by a Pullout Method. Curr Biol. 14(7):625-629.
Schmid M, Sieber R, Zimmermann Y-S, Vorburger C. 2012. Development, specificity and sublethal effects of symbiont-conferred resistance to parasitoids in aphids. Functional Ecology. 26(1):207-215.
Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 27(1367-4811).
Schmieder R, Lim YW, Edwards R. 2012. Identification and removal of ribosomal RNA sequences from metatranscriptomes. Bioinformatics. 28(3):433-435.
Schmitz A, Anselme C, Ravallec M, Rebuf C, Simon J-C, Gatti J-L, Poirié M. 2012. The Cellular Immune Response of the Pea Aphid to Foreign Intrusion and Symbiotic Challenge. PLoS ONE. 7(7):e42114.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30(14):2068-2069.
Shin M, Cha N, Koranteng F, Cho B, Shim J. 2020. Subpopulation of Macrophage-Like Plasmatocytes Attenuates Systemic Growth via JAK/STAT in the *Drosophila* Fat Body. Frontiers in Immunology. 11:63.
Silla-Martínez JM, Capella-Gutiérrez S, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25(15):1972-1973.
Smith AH, Łukasik P, O'Connor MP, Lee A, Mayo G, Drott MT, Doll S, Tuttle R, Disciullo RA, Messina A et al. 2015. Patterns, causes and consequences of defensive microbiome dynamics across multiple scales. Molecular Ecology. 24(5):1135-1149.
Smith TE, Moran NA. 2020. Coordination of host and symbiont gene expression reveals a metabolic tug-of-war between aphids and *Buchnera*. Proceedings of the National Academy of Sciences. 117(4):2113.
Soneson C, Love M, Robinson M. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; peer review: 2 approved]. F1000Research. 4(1521).

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9):1312-1313.
Tao H, Bausch C, Richmond C, Blattner FR, Conway T. 1999. Functional Genomics: Expression Analysis *Escherichia coli* Growing on Minimal and Rich Media. J Bacteriol. 181(20):6425.
van der Wilk F, Dullemans AM, Verbeek M, van den Heuvel JFJM. 1999. Isolation and Characterization of APSE-1, a Bacteriophage Infecting the Secondary Endosymbiont of Acyrthosiphon pisum. Virology. 262(1):104-113.
Vorburger C, Gouskov A. 2011. Only helpful when required: a longevity cost of harbouring defensive symbionts. Journal of Evolutionary Biology. 24(7):1611-1617.
Vorburger C, Perlman SJ. 2018. The role of defensive symbionts in host–parasite coevolution. Biological Reviews. 93(4):1747-1764.
Vorburger C, Rouchet R. 2016. Are aphid parasitoids locally adapted to the prevalence of defensive symbionts in their hosts? Bmc Evol Biol. 16(1):271.
Vorburger C, Sandrock C, Gouskov A, Castañeda LE, Ferrari J. 2009. Genotypic variation and the role of defensive endosymbionts in an all-parthenogenetic host-parasitoid interaction. Evolution. 63(6):1439-1450.
Wagner SM, Martinez AJ, Ruan Y-M, Kim KL, Lenhart PA, Dehnel AC, Oliver KM, White JA. 2015. Facultative endosymbionts mediate dietary breadth in a polyphagous herbivore. Functional Ecology. 29(11):1402-1410.
Weldon SR, Strand MR, Oliver KM. 2013. Phage loss and the breakdown of a defensive symbiosis in aphids. P Roy Soc B-Biol Sci. 280(1751).
Wenger JA, Cassone BJ, Legeai F, Johnston JS, Bansal R, Yates AD, Coates BS, Pavinato VAC, Michel A. 2017. Whole genome sequence of the soybean aphid, *Aphis glycines*. Insect Biochemistry and Molecular Biology.102917.
Whitaker D, Christman M. 2014. Clustsig: Significant Cluster Analysis. R package version 1.1.
Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA. 2003. Consequences of reductive evolution for gene expression in an obligate endosymbiont. Molecular Microbiology. 48(6):1491-1500.
Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15(3):R46.
Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology. 16(5):284-287.
Zanet J, Stramer B, Millard T, Martin P, Payre F, Plaza S. 2009. Fascin is required for blood cell migration during *Drosophila* embryogenesis. Development. 136(15):2557.
Zhang B, Horvath S. 2005. A general framework for weighted co-expression network analysis. Appl Genet Mol Biol. 4(Article 17).
Zhou A, He Z, Redding-Johanson AM, Mukhopadhyay A, Hemme CL, Joachimiak MP, Luo F, Deng Y, Bender KS, He Q et al. 2010. Hydrogen peroxide-induced oxidative stress responses in *Desulfovibrio vulgaris* Hildenborough. Environ Microbiol. 12(10):2645-2657.
Zhou X, Liao W-J, Liao J-M, Liao P, Lu H. 2015. Ribosomal proteins: functions beyond the ribosome. J Mol Cell Biol. 7(2):92-104.

Supplementary information



Supplementary Figure 1 – Cost of *H. defensa* strains varies. Impact of infection with different *H. defensa* strains on offspring production (cost) and parasitation rate (benefit) of aphid clone A06-407. Adapted from Cayetano et al. (2015). Aphid clone uninfected (black) or infected (infecting strains H15 (blue), H402 (orange), H76 (grey) or H85 (red)) by *H. defensa*. Error bars depict standard error. Data adapted from Cayetano et al. (2015).



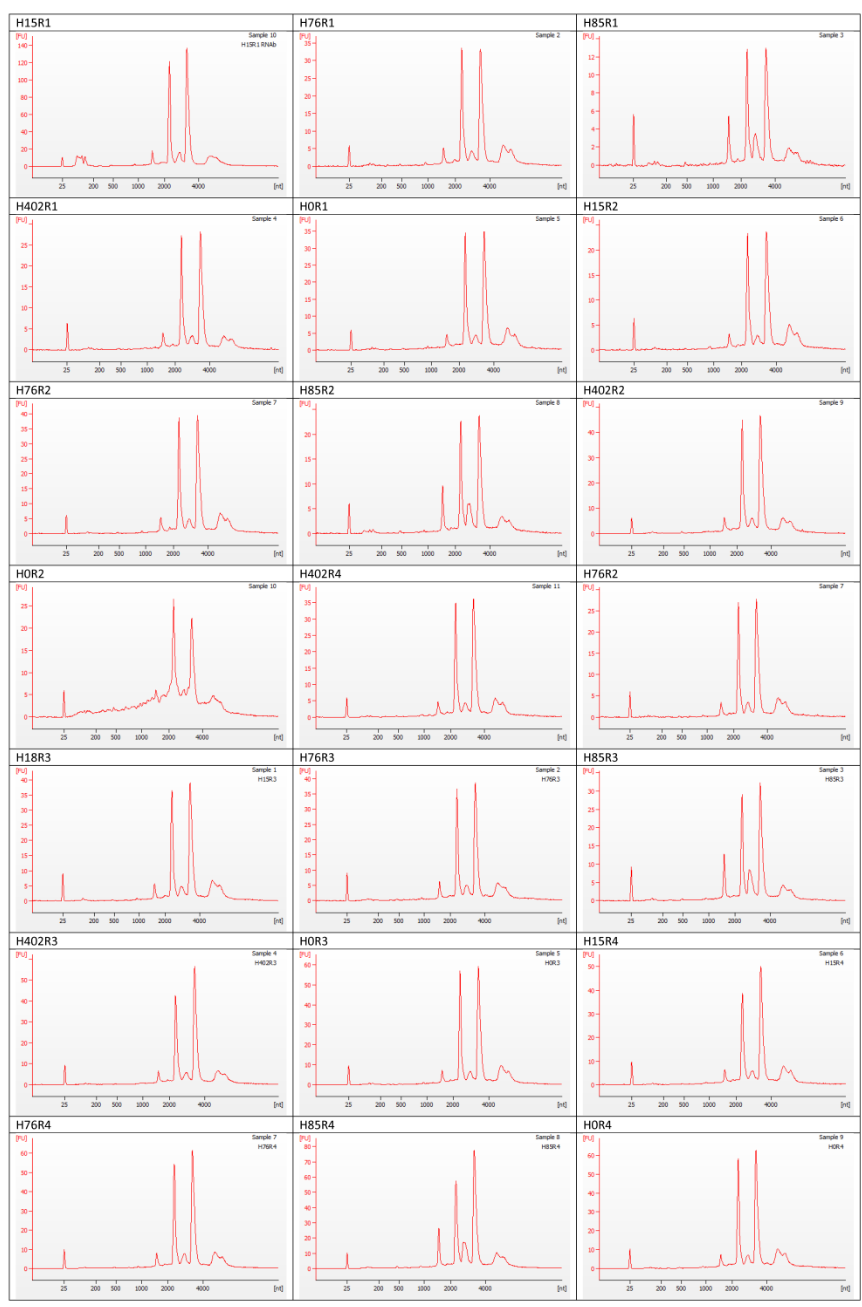
Supplementary Figure 2 – Library H15R1 is an outlier. Library H15R1 is located at the lower right corner of a PCA of the normalised and variance stabilisation transformed read count of all aphid genes. Aphid hosts were uninfected (H0, black) or infected (infecting strains H15 (blue), H402 (orange), H76 (grey) or H85 (red)) by *H. defensa*.

			Hamiltonella defensa												
		# genes	ApHdef-H10	ApHdef-H1	ApHdef-H11	ApHdef-H8	ApHdef-H7	ApHdef-H12	ApHdef-H9	ApHdef-H4	ApHdef-H2	ApHdef-H5	ApHdef-H3	ApHdef-H6	ApHdef-H13
			136	60	240	82	48	21	46	44	176	46	20	171	152
Aphis fabae	ApHdef-A2	141	0.08	0.36	-0.60	-0.58	0.77	-0.82	0.58	-0.39	0.33	-0.26	0.03	0.64	-0.81
	ApHdef-A4	243	-0.46	0.39	-0.91	-0.16	0.81	-0.49	0.08	-0.62	0.74	0.24	0.01	0.92	-0.73
	ApHdef-A5	99	-0.36	-0.02	-0.59	0.10	0.57	-0.24	-0.00	-0.68	0.45	0.34	-0.33	0.61	-0.32
	ApHdef-A7	383	0.46	-0.27	0.85	0.06	-0.76	0.43	-0.02	0.65	-0.70	-0.31	0.09	-0.86	0.62
	ApHdef-A6	22	-0.01	-0.69	0.01	0.36	0.33	-0.16	0.04	-0.69	-0.13	0.32	-0.85	-0.00	0.31
	ApHdef-A10	42	-0.46	0.30	-0.38	0.14	-0.06	0.42	-0.27	-0.05	0.45	0.25	0.21	0.37	-0.18
	ApHdef-A3	22	-0.90	0.42	-0.68	0.44	0.15	0.40	-0.71	-0.23	0.85	0.66	0.30	0.65	-0.20
	ApHdef-A1	42	0.73	0.41	0.23	-0.89	-0.13	-0.39	0.82	0.52	-0.40	-0.90	0.44	-0.18	-0.48
	ApHdef-A8	97	-0.09	-0.23	0.23	0.31	-0.39	0.56	-0.23	0.06	-0.12	0.16	-0.14	-0.24	0.39
	ApHdef-A9	121	-0.05	-0.06	-0.12	-0.01	0.08	0.06	0.09	-0.18	0.07	0.03	-0.16	0.13	-0.07
	ApHdef-A11	109	0.46	0.08	0.58	-0.27	-0.55	0.20	0.19	0.70	-0.49	-0.50	0.35	-0.58	0.20

Supplementary Figure 3 – Correlation of aphid and *H. defensa* gene modules. Pearson correlation coefficient between eigengenes of aphid and *H. defensa* modules. Coloured: Correlation has a p-value of <0.01, red indicates positive and blue indicates negative correlation.

			Hamiltonella defensa										
			BapHdef-H6	BapHdef-H8	BapHdef-H7	BapHdef-H2	BapHdef-H1	BapHdef-H3	BapHdef-H4	BapHdef-H10	BapHdef-H5	BapHdef-H11	BapHdef-H9
		# genes	60	136	240	48	21	46	44	176	171	82	152
Buchnera aphidicola	BapHdef-B2	11	0.11	-0.15	-0.24	-0.61	-0.15	-0.27	-0.20	0.25	0.35	0.31	-0.32
	BapHdef-B3	22	0.41	-0.36	0.18	-0.54	-0.18	0.21	-0.56	-0.28	0.53	-0.02	-0.55
	BapHdef-B5	11	-0.05	0.06	0.15	-0.75	-0.33	-0.14	-0.40	0.04	0.14	0.04	-0.14
	BapHdef-B1	82	-0.54	0.54	0.13	0.46	0.01	-0.12	0.38	0.12	-0.71	-0.18	0.71
	BapHdef-B4	42	0.73	-0.74	-0.24	-0.28	0.20	0.24	-0.26	-0.19	0.85	0.19	-0.85

Supplementary Figure 4 – Correlation of *B. aphidicola* and *H. defensa* gene modules. Pearson correlation coefficient between eigengenes of *B. aphidicola* and *H. defensa* modules. Coloured: Correlation has a p-value of <0.01, red indicates positive and blue indicates negative correlation.



Supplementary Figure 5 – High RNA integrity after total RNA extraction. Bioanalyzer 2100 electopherograms of all libraries in batch A and B. Comparison with Schroeder et al. (2006) implied a RIN of 8 for sample H0R2 and 9-10 for all others. Ribosomal RNA peaks of the different organisms are visible. Consider that a double 18S rRNA peak is expected for several insect species.

Supplementary Table 1 – Number of reads per organism. Number of reads in each library ('reads processed') and number and fraction of reads mapped to each organism by Kallisto. Column 'tmt' indicates treatment, i.e. uninfected (H0) or *H. defensa*-infected aphid hosts (infecting strains H15, H402, H76 or H85). 'Batch' indicates which libraries were grouped during RNA extraction and library preparation. The library that was excluded from analysis is marked with **. Read counts marked with * are most likely a result of index hopping.

Library	Batch	Reads processed	Number reads mapped			Fraction reads mapped		
			Hamiltonella defensa	*Buchnera aphidicola*	*Aphis fabae*	*Hamiltonella defensa*	*Buchnera aphidicola*	*Aphis fabae*
H0R1	A	34'348'692	*2355	3'203'120	21'832'963	0.0001	0.0933	0.6356
H0R2	A	34'040'040	*2568	3'153'843	22'046'894	0.0001	0.0927	0.6477
H0R3	B	34'222'139	*2511	3'189'260	22'340'889	0.0001	0.0932	0.6528
H0R4	B	35'200'355	*2492	3'158'555	23'578'885	0.0001	0.0897	0.6698
H15R1**	A	45'549'412	51'020	567'124	4'027'516	0.0011	0.0125	0.0884
H15R2	A	34'480'419	238'075	2'892'547	22'388'755	0.0069	0.0839	0.6493
H15R3	B	44'865'587	288'936	2'350'524	20'068'264	0.0064	0.0524	0.4473
H15R4	B	37'720'691	175'276	2'956'034	23'086'301	0.0046	0.0784	0.612
H402R1	A	32'394'543	438'745	3'053'013	20'631'605	0.0135	0.0942	0.6369
H402R2	A	32'275'213	426'496	2'899'940	20'767'985	0.0132	0.0899	0.6435
H402R3	B	32'403'389	534'593	2'581'333	21'390'798	0.0165	0.0797	0.6601
H402R4	B	34'966'568	539'990	2'935'467	23'320'347	0.0154	0.084	0.6669
H76R1	A	33'847'007	488'334	2'508'009	22'685'478	0.0144	0.0741	0.6702
H76R2	A	34'171'583	512'709	2'927'296	21'797'424	0.015	0.0857	0.6379
H76R3	B	33'061'665	395'952	3'363'776	20'997'915	0.012	0.1017	0.6351
H76R4	B	32'941'095	426'335	2'897'272	21'234'955	0.0129	0.088	0.6446
H85R1	A	33'680'986	4'537'843	2'087'672	17'343'683	0.1347	0.062	0.5149
H85R2	A	35'151'217	4'703'052	2'223'354	17'819'803	0.1338	0.0633	0.5069
H85R3	B	35'403'188	4'395'559	2'588'271	18'526'290	0.1242	0.0731	0.5233
H85R4	B	35'002'152	3'983'798	2'587'781	18'499'662	0.1138	0.0739	0.5285

Supplementary Table 2 – Read assignment to taxa. Fraction of reads assigned to different taxa by Kraken. Fractions were averaged over all libraries of each treatment, except for treatment H15 for which the strongly contaminated library H15R1 was listed separately from the other three libraries (H15*). Aphid hosts were uninfected (H0) or infected (infecting strains H15, H402, H76 or H85) by *H. defensa*. As this table contains only taxa of specific interest, the read fractions are not expected to add up to 100%. Fractions below 0.01 were not extracted from the Kraken output, which is indicated by "<0.01".

Kraken read classification	H0	H15*	H15R1	H76	H85	H402
Protostomia	53.32	45.54	8.05	52.32	41.79	52.92
Deuterostomia	0.14	1.98	14	0.15	0.13	0.1
Proteobacteria	6.94	8.09	12.16	8.01	21.59	8.16
Buchnera	6.38	4.64	1.07	5.82	4.46	5.9
Hamiltonella	<0.01	0.72	0.08	1.63	16.38	1.87
Escherichia	0.02	0.25	1.03	0.03	0.05	<0.01
Terrabacteria	0.12	2.78	17.18	0.12	0.14	0.07
Viruses	<0.01	0.11	<0.01	0.5	2.02	0.43
Bacteriophage APSE	<0.01	0.07	<0.01	0.38	1.64	0.35

Supplementary Table 3 – Completeness of assembly. BUSCO analysis of the aphid transcripts and prokaryote genes. BUSCO scores were calculated in relation to the reference databases arthropoda, insecta and proteobacteria.

	Aphis fabae		*Buchnera aphidicola*	*Hamiltonella defensa* consensus of H15 and H85
BUSCO database	arthropoda	insecta	proteobacteria	proteobacteria
Total BUSCO groups searched	1066	1658		221
% Complete BUSCOs	96.5	93.1	74.2	92.3
% Complete and single-copy BUSCOs	53.2	50.4	74.2	92.3
% Complete and duplicated BUSCOs	43.3	42.7	0.0	0.0
% Fragmented BUSCOs	0.8	2.6	1.4	3.2
% Missing BUSCOs	2.7	4.3	24.4	4.5

Supplementary Table 4 – Overview over the aphid's differential expression. Aphids infected with *H. defensa* (H15, H76, H85 or H402) were compared to aphids not infected by *H. defensa* (H0). In a reduced model containing only libraries of treatment H15 and H85, differential gene expression of aphids in presence of these two closely related *H. defensa* strains was analysed. For each pairwise comparison, number of differentially expressed genes (DEG), median fold changes of upregulated and downregulated genes as well as maximum fold change of upregulated and downregulated genes are indicated. No GO-terms were found enriched among the DEG.

Pairwise comparison	Number DEG	Median fold change of upregu-lated genes	Median fold change of down-regulated genes	Maximum fold change of upregulated genes	Maximum fold change of downregulated genes	GO terms enriched in DEG
H15 vs H0	42	0.88	-0.85	1.72 (uncharacterized protein)	-3.36 (G patch domain-containing protein 11)	none
H76 vs H0	11	1.34	-0.91	4.18 (uncharacterized protein)	-5.1 (uncharacterized protein)	none
H85 vs H0	19	0.9	-1.05	1.78 (uncharacterized protein)	-4.34 (uncharacterized protein)	none
H402 vs H0	32	1.88	-1.85	1.28 (uncharacterized protein)	-3.36 (uncharacterized protein)	none
H15 vs H85	6	1.56	-1.37	1.71 (ubiquitin-related modifier 1)	-3.79 (uncharacterized protein)	none

Supplementary Table 5 – Differential expression of aphid genes (full model). Results of differential expression analysis comparing gene expression of aphids infected with *H. defensa* (H15, H76, H85 or H402) to aphids not infected by *H. defensa* (H0). P-values adjusted for multiple testing (padj) <0.01 and absolute log2 fold changes(L2FC) >0.5 are indicated by coloured backgrounds. *This table is provided in the electronic supplementary material.*

Supplementary Table 6 – Differential expression of aphid genes (reduced model). Results of differential expression analysis comparing gene expression of aphids infected with *H. defensa* H15 to aphids infected with *H. defensa* H85. P-values adjusted for multiple testing (padj) <0.01 and absolute log2 fold changes(L2FC) >0.5 are indicated by coloured backgrounds. *This table is provided in the electronic supplementary material.*

Supplementary Table 7 – Differential expression of *H. defensa* genes (full model). Results of differential expression analysis comparing gene expression of *H. defensa* strains H15, H76 or H402 to *H. defensa* strain H85 (full model with all strains). P-values adjusted for multiple testing (padj) <0.01 and absolute log2 fold changes(L2FC) >0.5 are indicated by coloured backgrounds. *This table is provided in the electronic supplementary material.*

Supplementary Table 8 – Differential expression of *B. aphidicola* genes. Results of differential expression analysis comparing gene expression of *B. aphidicola* in aphid hosts infected with *H. defensa* (H15, H76, H85 or H402) to uninfected aphid hosts (H0). P-values adjusted for multiple testing (padj) <0.01 and absolute log2 fold changes(L2FC) >0.5 are indicated by coloured backgrounds. *This table is provided in the electronic supplementary material.*

Supplementary Table 9 – Correlation of aphid gene expression. (A) WGCNA analysis with aphid and *H. defensa* genes. Table includes correlation coefficient and p-values of aphid gene modules with *H. defensa* titre. Coloured: Correlation has a p-value of <0.01. If no GO-terms are enriched in a module, the value in 'associated GO-terms' is set to 'NA'. (B) Correlation analysis with aphid and *H. defensa* genes. *This table is provided in the electronic supplementary material*.

Supplementary Table 10 – Correlation of *H. defensa* gene expression. (A) WGCNA analysis with aphid and *H. defensa* genes. Table includes correlation coefficient and p-values of *H. defensa* gene modules with *H. defensa* titre. Coloured: Correlation has a p-value of <0.01. If no GO-terms are enriched in a module, the value in 'associated GO-terms' is set to 'NA'. (B) Correlation analysis with aphid and *H. defensa* genes. (C) Correlation analysis with *H. defensa* and *B. aphidicola* genes. *This table is provided in the electronic supplementary material.*

Supplementary Table 11 – Correlation of *B. aphidicola* gene expression. (A) WGCNA analysis with *B. aphidicola* and *H. defensa* genes. Table includes correlation coefficient and p-values of *B. aphidicola* gene modules with *H. defensa* titre. Coloured: Correlation has a p-value of <0.01. If no GO-terms or KEGG pathways are enriched in a module, the values in 'associated GO-terms' and 'associated KEGG pathways' is set to 'NA'. (B) Correlation analysis with *B. aphidicola* and *H. defensa* genes. *This table is provided in the electronic supplementary material*.

Supplementary Table 12 – Origin of aphid clones. Collection date, site and host plant for the aphid clones used as donors and recipients during the transfections that created the infected A06-407 sublines.

Aphid clone	Collection date	Collection site	Host plant	*H. defensa* isolate
A06-15	08.05.2006	Ressora, Italy	*Vicia faba*	H15
A06-76	17.05.2006	La Grande Motte, France	*Chenopodium album*	H76
A06-85	17.05.2006	Grimaud, France	*Chenopodium album*	H85
A06-402	01.07.2006	St. Margrethen SG, Switzerland	*Chenopodium album*	H402
A06-407	01.07.2006	St. Margrethen SG, Switzerland	*Chenopodium album*	naturally uninfected

Supplementary Table 13 – Accession numbers of *H. defensa* assemblies. Protein fasta files from genome assemblies of *H. defensa* were downloaded from NCBI and provided to PROKKA for gene prediction and to BUSCO for phylogenetic analyses.

Strain	Host organism	RefSeq/GenBank accession	Name
T5A	*Acyrthosiphon pisum*	GCF_000021705.1	ASM2170v1
A2C	*Acyrthosiphon pisum*	GCF_002777195.1	ASM277719v1
AS3	*Acyrthosiphon pisum*	GCF_002777215.1	ASM277721v1
ZA17	*Acyrthosiphon pisum*	GCF_002777235.1	ASM277723v1
NY26	*Acyrthosiphon pisum*	GCF_002777295.1	ASM277729v1
MED	*Bemisia tabaci*	GCA_000258345.2	ASM25834v2
MEAM	*Bemisia tabaci*	GCF_002285855.1	ASM228585v1

Supplementary Table 14 – Accession numbers of contaminant bacterial genomes. Coding DNA sequences (CDS) from genome assemblies of the most frequent contaminant bacteria were downloaded from NCBI and provided to Kallisto during mapping.

Organism	RefSeq/GenBank accession	Name
Bacillus subtilis	GCF_000009045.1	ASM904v1
Escherichia coli	GCF_000005845.2	ASM584v2
Klebsiella pneumoniae	GCF_000240185.1	ASM24018v2
Providencia stuartii	GCF_000259175.1	ASM25917v1
Rahnella aquatilis	GCF_003573465.1	ASM357346v1

Chapter 2

Horizontal transmission of the heritable protective endosymbiont *Hamiltonella defensa* depends on titre and haplotype

Heidi Kaech[1,2], Christoph Vorburger[1,2]

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[2]D-USYS. Department of Environmental Systems Science, ETH Zürich, Switzerland

Published: 2021; Frontiers in Microbiology, Microbial Symbioses 11: 3577-3587
https://doi.org/10.3389/fmicb.2020.628755

Abstract

Secondary endosymbionts of aphids have an important ecological and evolutionary impact on their host, as they provide resistance to natural enemies but also reduce the host's lifespan and reproduction. While secondary symbionts of aphids are faithfully transmitted from mother to offspring, they also have some capacity to be transmitted horizontally between aphids. Here we explore whether eleven isolates from three haplotypes of the secondary endosymbiont *Hamiltonella defensa* differ in their capacity for horizontal transmission. These isolates vary in the protection they provide against parasitoid wasps as well as the costs they inflict on their host, *Aphis fabae*. We simulated natural horizontal transmission through parasitoid wasps by stabbing aphids with a thin needle and assessed horizontal transmission success of the isolates from one shared donor clone into three different recipient clones. Specifically, we asked whether potentially costly isolates reaching high cell densities in aphid hosts are more readily transmitted through this route. This hypothesis was only partially supported. While transmissibility increased with titre for isolates from two haplotypes, isolates of *H. defensa* haplotype 1 were transmitted with greater frequency than isolates of other haplotypes with comparable titres. Thus, it is not sufficient to be merely frequent – endosymbionts might have to evolve specific adaptations to transmit effectively between hosts.

Keywords: *Aphis fabae*, *Hamiltonella defensa*, horizontal transmission, symbiont, titre

Introduction

Microbial symbionts are ubiquitous in insects. They often influence ecologically relevant traits and therefore the ecological niche of their hosts. For aphids, survival on a diet of amino-acid-poor plant sap is only possible because of a symbiosis with the γ-proteobacterium *Buchnera aphidicola,* which lives in specialised tissues of their body (Baumann 1995). Together, aphid and endosymbiont produce all the essential amino acids lacking from the aphid's diet (Hansen and Moran 2011). On their own, neither of the two organisms can survive, making this symbiosis a primary endosymbiosis. In addition to *B. aphidicola,* aphids can also harbour other bacterial endosymbionts. While these provide benefits to the aphid, such as protection against natural enemies (Ferrari et al. 2004; Oliver et al. 2003) or improved resistance to heat stress (Chen et al. 2000; Russell and Moran 2006), they are not essential for the host's survival. They are therefore referred to as secondary endosymbionts.

To be maintained in a host population, symbionts must transmit vertically from mother to offspring, horizontally between individuals, or both. Transmission mode is an evolutionary continuum that can change over time and impacts the relationship between host and symbiont (Ebert 2013). At one end of the spectrum are the symbionts with strict and faithful vertical maternal transmission. Their dispersal depends completely on their host's reproduction. Ideally, they limit the amount of energy that they extract from their host to the absolute minimum so that the host can produce a maximal number of offspring, that lead to the symbiont's dispersal. Thus, the intimate association with their host's reproduction drives vertically transmitted endosymbionts towards avirulence (Bull et al. 1991). At the other end of the spectrum, there are purely horizontally transmitted symbionts. As their dispersal does not depend on their host's reproduction, they tend to extract enough energy from their host to maximise their horizontal transmission success – for example by increasing the number of infectious particles that can spread (Ewald 1983). As a result of their lifestyle, purely horizontally transmitted symbionts tend to be more virulent than vertically transmitted symbionts (Fisher et al. 2017).

Transmission mode has strong implications for the ecology and coevolution of both host and symbiont (Chrostek et al. 2017). For example, competition between different horizontally transmitted symbiont strains can, but does not have to, lead to greater host exploitation through an escalation of virulence as both competitors try to obtain a greater share of the host's resources (Frank 1992; McLean et al. 2018; Nowak and May 1994). Horizontal

transmission via vectors is thought to be particularly prone to lead to high virulence in the host (Ewald 1983) and costly endosymbionts can use occasional horizontal transmission in combination with vertical transmission to be maintained in competition with less virulent and purely vertically transmitted strains (Lipsitch et al. 1995).

The primary endosymbiont of aphids, *B. aphidicola*, relies purely on vertical transmission and lives intracellularly (Baumann 1995). The situation of secondary endosymbionts like *Hamiltonella defensa* and *Regiella insecticola* (Moran et al. 2005c) is more ambiguous. On one hand, secondary endosymbionts transmit vertically with nearly 100% efficiency in the lab (Darby and Douglas 2003; Fukatsu et al. 2000; Vorburger et al. 2017), such that they are hardly ever lost from laboratory cultures. On the other hand, they have the potential to transmit horizontally, as they do not only occur intracellularly but also in the hemolymph (Fukatsu et al. 2000). In fact, occasional horizontal transmission is necessary to explain the strain distribution of secondary endosymbionts across aphids (Henry et al. 2013; Russell et al. 2003; Sandström et al. 2001). In nature, horizontal transmission of aphid facultative endosymbionts could occur through the sting of parasitoid wasps (Gehrer and Vorburger 2012; Heath et al. 1999; Peccoud et al. 2014), through sexual transmission (Moran and Dunbar 2006), or – based on evidence from other insects – via plant tissues and surface contamination (Caspi-Fluger et al. 2012; Darby and Douglas 2003).

In this work, we investigated the horizontal transmission potential of 11 different isolates of the secondary endosymbiont *H. defensa* in black bean aphids (*Aphis fabae*). The symbiont provides protection against parasitoid wasps, but the strength of this protection varies greatly between different isolates (Cayetano et al. 2015). In addition, there are large differences in the cost and the titre that different isolates reach within their host (Cayetano et al. 2015). Interestingly, strength of protection and cost to the host are negatively, rather than positively, related across different isolates. This raises the question of how highly virulent isolates that provide limited benefits can persist in aphid populations.

We hypothesized that highly protective, avirulent isolates of *H. defensa* rely primarily on vertical transmission, whereas costly, less protective isolates also depend on horizontal transmission. Costly isolates may extract more resources from their host to increase their abundance in the host's hemolymph (Chong and Moran 2016b), which in turn may increase their chance to be successfully transmitted from one aphid to another by a parasitoid's ovipositor. To test this hypothesis, we simulated horizontal transmission events using fine

needles and correlated the titre of different *H. defensa* isolates with their horizontal transmission success.

Methods

Aphid clones and *H. defensa* isolates

For the horizontal transmission assay, we used 12 sublines of the *A. fabae* clone A06-407 as hemolymph donors. Clone A06-407 was originally free from secondary endosymbionts ('407H0') and had been microinjected with 11 different *H. defensa* isolates from other *A. fabae* clones between 2008 and 2012 (Supplementary Table 1) to form sublines 407H15 to 407HAf6. Date of creation of sublines and collection details of aphid clones are provided in Supplementary Table 1 and Supplementary Table 2, respectively. Based on partial sequences of two bacterial housekeeping genes, *murE* and *accD,* the *H. defensa* isolates can be grouped into three haplotypes: Haplotype 1 comprising H76 and H101, haplotype 2 comprising H9, H28, H30, H323, H343, H402 and AF6, and haplotype 3 comprising H15 and H85 (Cayetano et al. 2015) (Supplementary Table 1). The division into these three haplotypes has been confirmed by sequencing additional genes (Youn Henry, personal communication). Costs and benefits of different *H. defensa* isolates were determined three years prior to this experiment by Cayetano et al. (2015). Both costs and benefits varied strongly between isolates, with highly protective isolate H76 causing no detectable reduction in lifetime offspring production and the less beneficial isolate H85 strongly reducing the amount of offspring. Costs of the other strains lay in between these two extremes (Supplementary Figure 1).

Hemolymph was transferred from each donor to three recipient *A. fabae* clones (A06-37, A06-405 and A06-407). The recipient clones were collected in summer 2006 in Europe (Supplementary Table 2) and are naturally free of any known secondary endosymbionts (Vorburger et al. 2009). Recipient names are abbreviated as 37H0, 405H0 and 407H0, with "H0" indicating the absence of the secondary endosymbiont *H. defensa*. We used three recipients to assess whether different genotypes differ in how readily they accept the *H. defensa* isolates, and we included clone A06-407 as a recipient to test whether *H. defensa* isolates have adapted to this clone since their introduction and are therefore more easily (re-)introduced to it. A fourth *H. defensa*-free clone, which is visibly distinguishable from all others due to a colour mutation, A08-28^{H-} (Supplementary Table 2), was included in the assays to check for potential horizontal transmission via plants (see Experimental Procedures below). Since collection or creation, all aphid clones and sublines were maintained in a clone

bank on broad bean (*Vicia faba)* seedlings and under conditions that ensured clonal reproduction (18-20°C, 16 h photoperiod).

Experimental procedures

Subadult aphids (nymphs) of all three recipient clones were stabbed with fine pins contaminated by hemolymph from each of the 11 *H. defensa*-infected sublines of clone A06-407 as well as hemolymph from 407H0 as a symbiont-free control. All 36 donor-recipient combinations were replicated 5 times, with each replicate consisting of a batch of 10 stabbed nymphs. Transmission success was determined by testing the offspring of stabbed individuals for the presence of *H. defensa*. Therefore, we only counted horizontal transmissions as successful if they led to new heritable infections.

To produce enough aphids for the experiment, five adults of each aphid clone or subline were split off from the laboratory stocks and bred up on *V. faba* seedlings for two generations. Adults used to found the third generation were allowed to reproduce on new plants for 24 h, then they were frozen at -20°C and used to confirm the presence/absence of *H. defensa* and the correct *H. defensa* haplotype at the start of the experiment (see Molecular Methods below).

To produce the aphids acting as hemolymph donors, 10 adults of the second generation of each subline were placed on a new bean plant, allowed to reproduce for 24 h and then transferred to a next plant. This was repeated over five days to produce five consecutive batches of offspring. Offspring were reared for 15 days until they reached adulthood, at which point they were used as donors in the transmission assays.

To produce the aphids acting as hemolymph recipients, each of the three recipient clones was reared on 12 separate plants for two generations. We then took six adults from each plant, placed them on a new plant to reproduce for 24 h, before moving them on to another new plant. Again, this was repeated over five days to produce five batches of offspring. Offspring were used as recipients in the transmission assays as 3-day old nymphs.

Manual transfection of hemolymph from donor to recipient took place over five consecutive days. For every combination of donor and recipient, one 15-day old donor aphid was used to infect ten 3-day old recipient nymphs. The donor was mounted with double-sided tape on a glass slide and stung with a fine stainless steel needle (Minutien pins, 0.1 mm diameter, Fine Science Tools GMBH), which was then inserted briefly into a recipient nymph. This double

stabbing procedure was repeated for each of the ten nymphs. Before use, the needles had been sanitized by soaking in 70% ethanol for five minutes, but they were not cleaned between successive stabs of the same donor-recipient clone combination. Aphids were under CO_2-anesthesia during the procedure. Except for the first day, the donors were frozen after use for later reference. On each day we collected a pool of three donors per subline, which was frozen at -20°C until DNA extraction and subsequent qPCR for estimation of *H. defensa* titres.

After transfections, the ten nymphs of each donor-recipient combination were transferred to a ventilated insect breeding dish (Ø 5 cm), which contained a broad bean leaf disc (Ø 4 cm) on 1% agar, and were maintained at 21°C and a 16 h photoperiod. To test whether rearing all 10 nymphs on the same leaf disc might allow between-nymph transmission of *H. defensa* via the leaf tissue, we also added two three-day-old nymphs of the *H. defensa*-free colour-mutant clone A08-28^{H-} as sentinels to each disc. The number of surviving recipients in each disc was scored six days after transfection, when aphids approached adulthood. Either four (batch 1) or three (batches 2-5) survivors were transferred individually to new leaf discs to be reared until their offspring reliably indicated status of infection. Preliminary experiments had established this to be the case at approximately 13-days old, a finding which is corroborated by literature (Chen and Purcell 1997). The two sentinel aphids of each leaf disc were raised together on a new leaf disc, reared for seven days and then frozen at -20°C until DNA extraction.

Three days after isolating the surviving recipients on separate leaf discs, the number of offspring produced by each survivor was noted (time point t1) and the offspring discarded. After another four days, the survivors were moved to new leaf discs and the offspring on the old leaf disc were counted (time point t2) and discarded. After reproducing for two days on the new leaf discs, the survivors were removed from the leaf disc and frozen at -20°C. After an additional five days, a pool of three offspring of each survivor was harvested for diagnostic PCR for *H. defensa* to test whether their stabbed mothers had acquired and passed on the symbiont. The three nymphs from each leaf disc were collected into collection microtubes of the DNeasy 96 Blood & Tissue kit (Qiagen) to which two sterile glass beads were added.

In the second batch, offspring of one A06-407 recipient exposed to hemolymph from donor 407H0 tested positive for *H. defensa*, which is only explicable by contamination. To identify its source, we tested all available H0 donors (batches 2-5) and the mothers of all A06-407 recipients for presence/absence of *H. defensa*. This revealed that the recipient culture was clean, but the donor culture of 407H0 had been contaminated with *H. defensa*-infected

individuals. Four of the 12 H0 donors used in batches 2-5 were found to have carried *H. defensa*. We discarded their data as well as all data from H0 donors in batch 1, since these donors had not been saved and could therefore not be checked retrospectively.

Molecular methods

DNA was extracted from aphids using a 'salting out' protocol (Sunnucks and Hales 1996) for checking the identity of aphids used to set up the experiment. For estimating *H. defensa* titres in donor lines and for assessing whether offspring produced by recipient aphids carried *H. defensa* we used the Qiagen DNeasy 96 Blood & Tissue kit (Qiagen AG, Hombrechtikon, Switzerland). Extraction success was verified by amplifying part of the 16S rRNA gene of *B. aphidicola*, the obligate endosymbiont present in all aphids, using specific primers. The presence/absence of *H. defensa* was also determined by diagnostic PCR with specific primers for the same gene. Primers and cycling conditions are detailed in Supplementary Table 3. Amplicons were visualized either by agarose gel electrophoresis or by capillary electrophoresis on a QIAxcel Advanced System (Qiagen AG, Hombrechtikon, Switzerland). To identify *H. defensa* haplotypes, we amplified *murE* and *accD* gene fragments (primers and cycling conditions in Supplementary Table 3) for Sanger sequencing by a commercial provider (GATC Biotech AG, Köln, Germany).

The density of *H. defensa* in the 11 donor lines was estimated from five replicate pools of three individuals (one from each batch), using TaqMan real-time quantitative PCR as described by Schmid et al. (2012). The ratio of *H. defensa*'s *dnaK* and the aphid's *EF1α* gene served as a proxy for *H. defensa* titre. The 20 µl reactions were run in triplicates on a LightCycler 480 (Roche) and gene copy number was estimated based on a standard curve produced with serial dilutions of a synthetic standard provided by Microsynth AG (Balgach, Switzerland). For analysis, we used the 2nd Derivative Absolute Quantification approach in LightCycler 480 SW 1.5.1. Two replicates were excluded because we failed to harvest three live adults for DNA extraction (batch 3 of 407H323 and batch 4 of 407H85).

Statistical analysis

Statistical analyses were performed in RStudio v1.2.5033 (RStudio Team 2020) and R v3.6.3 (R Core Team 2018) using the packages reshape2 v1.4.4 (Wickham 2007) and dplyr v0.8.5 (Wickham et al. 2020) for data wrangling, lme4 v1.1-23 (Bates et al. 2015) for linear mixed models and generalized linear mixed models, DHARMa v0.3.3.0 (Hartig 2020) for residual analysis and ggplot2 v3.3.2 (Wickham 2016) for producing figures. Post-hoc tests were

performed with multcomp v1.4-13 (Hothorn et al. 2008). Anova-like tables of random effect terms of linear mixed models using likelihood ratio tests were produced with lmerTest v3.1-2 (Kuznetsova et al. 2017).

To compare *H. defensa* titres among different isolates and haplotypes, the ratio of *H. defensa dnaK* to *A. fabae EF1α* copy numbers was log-transformed and used in a linear mixed model with haplotype as fixed and *H. defensa* isolate as random effect. Since haplotype did not significantly influence titre, we also ran a linear model with just *H. defensa* isolate as predictor. For further analyses, the titre of each isolate was averaged over all five batches. For models with transmission rate as response variable, we only included donors that carried *H. defensa* (i.e. excluding 407H0). We used a generalised linear mixed model (GLMM) with binomial errors, expressing transmission rate as aggregated binomial data juxtaposing number of infected to number of uninfected individuals per treatment and batch. We treated average titre, recipient clone and donor haplotype as fixed effects and included the interactions between recipient and titre and the interaction between recipient and haplotype. Donor identity and batch were treated as random intercepts. We assumed there to be no interaction of titre and haplotype. The data indicates that this is true for haplotype 2 and 3, and it is impossible to determine a reliable estimate of the slope for haplotype 1 since both isolates of this haplotype (H76 and H101) have near-identical transmission rates and titre. To allow the model to converge, we used the Bobyqa optimizer with a set maximum of 200’000 iterations. Uniformity of residuals was confirmed using the package DHARMa (Hartig 2020), which uses a simulation-based approach, and the model was checked for overdispersion.

To analyse survival of stabbed nymphs, we used a linear mixed effects model on the arcsine-square-root-transformed proportion of surviving nymphs. Average titre, recipient clone and donor haplotype (including *H. defensa*-free donors as a fourth ‘haplotype’) as well as the interaction between recipient and titre and recipient and haplotype were treated as fixed effects, and batch and donor were random intercepts.

The fecundity of survivors during the first week of their adult life was analysed using a GLMM with negative binomial errors. We excluded aphids that had not survived until time point t2 from the data set as their infection status could not be reliably assessed. Recipient clone and infection status (‘control’, donor not infected with *H. defensa*; ‘exposed’, donor infected with *H. defensa* but the infection was not transmitted to the survivor’s offspring; ‘infected’, the infection was transmitted to the survivor’s offspring) were treated as fixed

effects, and batch and donor as random intercepts. In an additional GLMM fecundity of successfully infected aphids was compared. Donor was treated as a fixed effect and batch as a random intercept.

Results

Of 1800 recipients (five times 10 nymphs for each of the 36 donor-recipient combinations) that were stabbed with hemolymph-contaminated needles, 1209 (67%) survived the transfection. Survival was different between batches (χ^2=18.28, df=1, p<0.001) but not between different isolates of *H. defensa* (χ^2=0.01, df=1, p=0.943), and did not depend on titre ($F_{1,7}$=0.08, p=0.786) or haplotype ($F_{3,9}$=0.12, p=0.943) of the *H. defensa* isolate that was transfected. Also non-significant were the effects of the recipient clone ($F_{2,147}$=1.05, p=0.353), the interaction between titre and recipient ($F_{2,147}$=2.16, p=0.119) and the interaction between recipient and haplotype ($F_{6,147}$=0.79, p=0.583). The significant influence of batch on survival was likely a result of improving skill and routine of the experimenter (Figure 1 A).

A maximum of four (in batch 1) or three (in batches 2-5) nymphs of each treatment that survived the transfection were kept until their offspring would reliably indicate infection with *H. defensa*. We worked with 479 survivors instead of 576, since we did not always have as many nymphs survive as desired (n=50), survivors died before their offspring would reliably indicate *H. defensa* infection (n=28), and we discarded recipients stung with 407H0 donors that were not confirmed to be free of contamination with *H. defensa* (n=19). On average, survivors produced 40.5±0.5 nymphs during their first week of reproduction. Fecundity in early adult life varied significantly among the three recipient clones (χ^2=36.98, df=2, p<0.001), with clone 37H0 producing more offspring (Figure 1 B), but did not depend on the recipient's infection status (χ^2=5.10, df=2, p=0.078). Thus the number of offspring produced in early adult life was similar between aphids that had acquired an infection with *H. defensa,* those that had rejected an infection and those that were not exposed to infectious hemolymph (Figure 1 B). Among aphids that had been successfully infected through the transfection, fecundity was not significantly influenced by the infecting isolate's identity (χ^2=10.85, df=10, p=0.37).

Horizontal transmission of *H. defensa* isolates from donor to recipients was considered successful if the symbiont managed to transmit to the recipient's offspring. Of the 479 survivors of the transfection, 458 had been stabbed with hemolymph of *H. defensa*-infected donors. Of these, 156 (34%) produced *H. defensa*-positive offspring. These are indeed

infections acquired from the stab and not infections acquired secondarily from other nymphs feeding on the same leaf disc, because none of the 177 sentinel aphids we tested were positive for *H. defensa*.

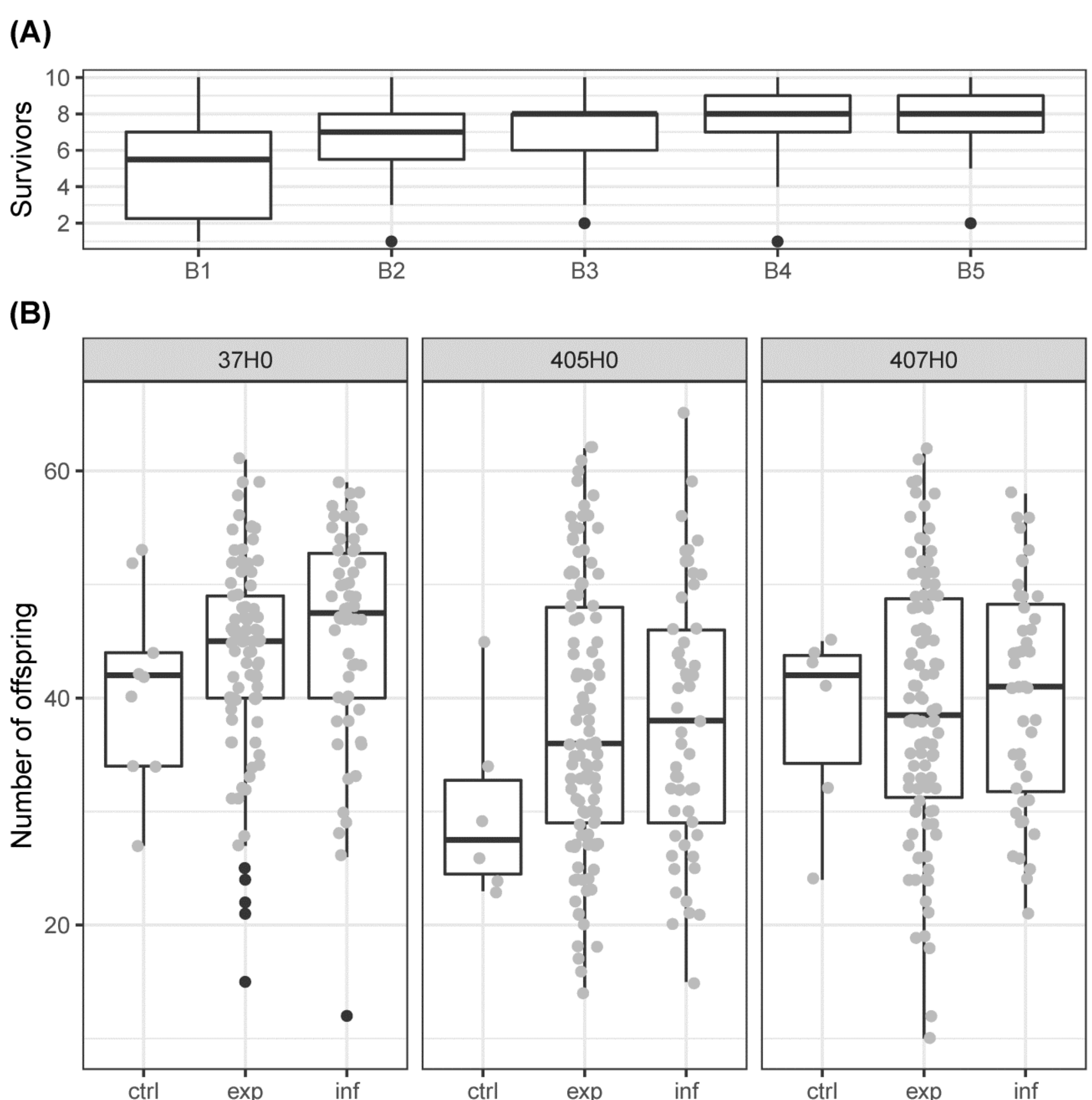


Figure 1 – (A) The number of nymphs out of ten that survived being stung with a hemolymph-contaminated needle increased while processing the five consecutive experimental batches B1 to B5. (B) Reproduction of young adults of clones 37H0, 405H0 or 407H0 during one week. Recipients had been stung with a needle contaminated with hemolymph from a donor without *H. defensa* (control, 'ctrl') or with hemolymph from a donor with *H. defensa*. Recipients that had been exposed to potentially infectious hemolymph could either reject the infection ('exp') or succumb to it ('inf'). Black dots indicate boxplot outliers.

Horizontal transmission success varied between different *H. defensa* isolates and batches, which was reflected in a highly significant random effect of *H. defensa* isolate and a significant effect of batch (Table 1, Figure 2). Both haplotype and titre significantly affected the transmission rate of an isolate (Table 1). Endosymbiont titre – expressed as the ratio of

H. defensa dnaK to *A. fabae EF1α* copy numbers – varied significantly among different *H. defensa* isolates ($F_{10,42}$=39.86, p<0.001) but not between haplotypes ($F_{2,8}$=0.38, p=0.697) (Supplementary Figure 2, Figure 2).

Table 1 – Results of a generalised linear mixed effects model for the transmission rate of different *H. defensa* isolates to different recipients (37H0, 405H0, 407H0). Model predictors were recipient, average titre that an isolate reached in the donor aphid and haplotype of the isolate (haplotypes 1, 2 and 3). The aphid subline acting as donor during horizontal transmission ('donor') and experimental batch were treated as a random effect.

	Effect	LR χ^2	df	p-value
Random:	donor	24.48	1	**<0.001**
	batch	4.29	1	**0.038**
Fixed:	titre of *H. defensa* isolate	8.93	1	**0.003**
	haplotype of *H. defensa* isolate	11.18	2	**0.003**
	recipient clone	3.52	2	0.172
	titre : recipient	8.81	2	**0.012**
	recipient : haplotype	10.26	4	**0.036**

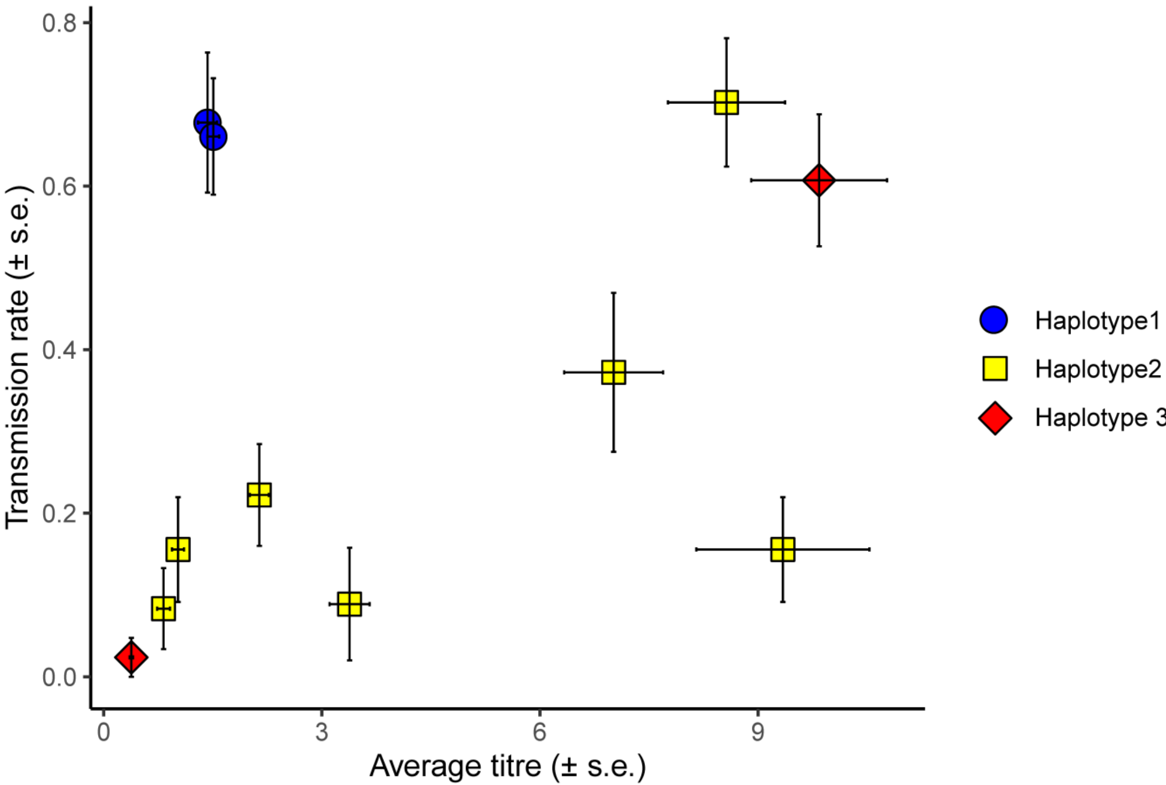


Figure 2 – Average titre of different *H. defensa* isolates in donor aphids plotted against the average transmission rate of the isolate. Transmission rate corresponds to the number of recipients out of three or four in which a *H. defensa* isolate successfully established after horizontal transmission, i.e. was propagated to the recipient's offspring. In this figure, transmission rate is averaged over three different recipients (37H0, 405H0 and 407H0) and five experimental batches. Error bars indicate the standard error and the combination of colours and symbols indicates the haplotype of the *H. defensa* isolate (blue circle = haplotype 1, yellow square = haplotype 2, red diamond = haplotype 3).

In general, *H. defensa* isolates that reached a high titre in the donor were more frequently transmitted (Figure 2 & 3): The seven isolates of haplotype 2 varied widely in their titres, and their horizontal transmission rates increased with titre. The two isolates of haplotype 3 reached the lowest and one of the highest titres, and the high-density isolate H85 was transmitted much more frequently than the low-density isolate H15. However, isolates H101 and H76 of haplotype 1, which both had comparably low titres, were much more successful at horizontal transmission than isolates from other haplotypes with similar titre (Figure 2).

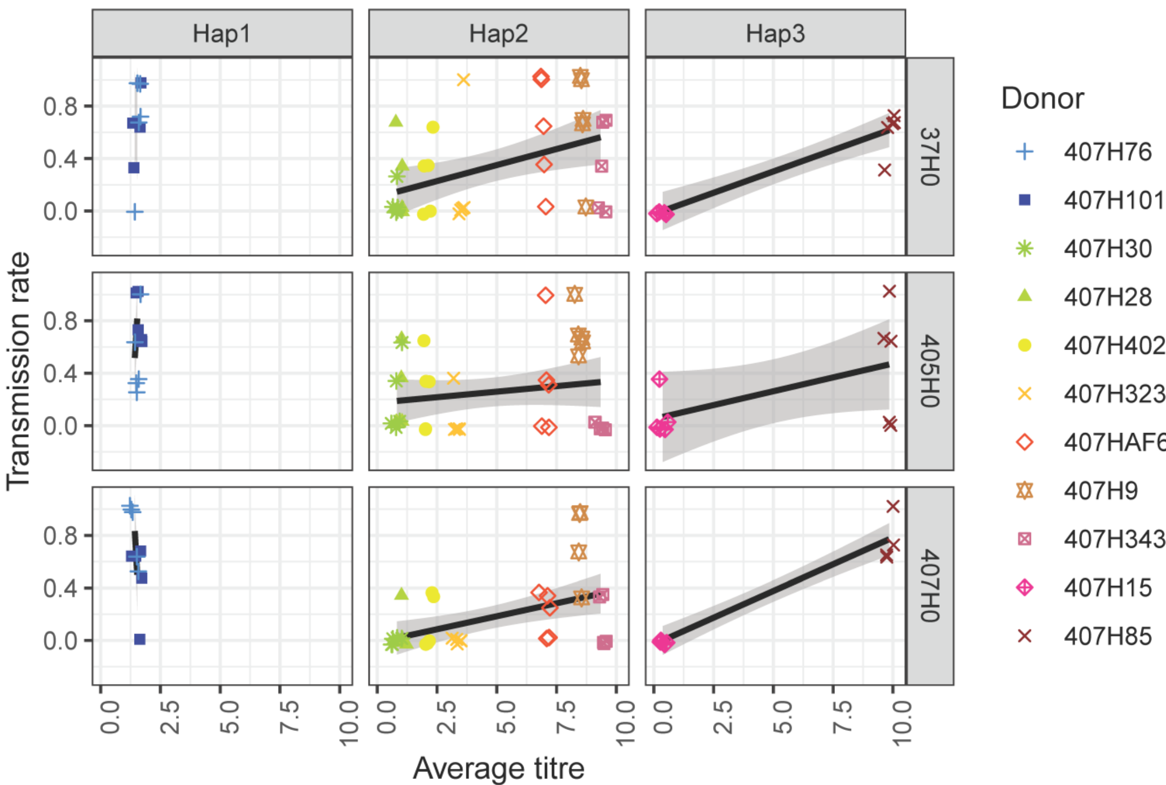


Figure 3 – Eleven *H. defensa* isolates were horizontally transfected from donor aphids (clone A06-407) to recipients (37H0, 405H0 and 407H0). Transmission rate corresponds to the number of recipients out of three or four in which a *H. defensa* isolate successfully established after horizontal transmission, i.e. was propagated to the recipient's offspring. The combination of colours and symbols indicate identity of the different isolates ('donor'). Transmission rate is plotted against average titre of the isolate on the x-axis. To allow that data points with similar coordinates can be discerned, a jitter of 0.1 has been applied to both the x- and the y-axis.

Recipient clones did not differ in how frequently they acquired *H. defensa* (Table 1). Hence there was no evidence that *H. defensa* established more readily in the same genetic background as it was maintained in for several years. However, since recipient 405H0 was less susceptible to infection with high-titre isolates than 407H0 and 37H0 (Figure 3), there was a significant interaction between titre and recipient (Table 1). Also, isolates of haplotype 2 transmitted somewhat less frequently to recipient 407H0 (Figure 3), leading to a marginally significant interaction between recipient and haplotype (Table 1). Recipients of the first

experimental batch showed rather poor and highly variable survival (Figure 1A). To verify that the horizontal transmission results were not overly influenced by this batch, we repeated the analysis with a reduced dataset only containing batches 2–5. The random batch effect was indeed no longer significant in this analysis, but all the significant fixed effect remained significant (Supplementary Table 4), and the estimates of horizontal transmission rates remained nearly unchanged (Supplementary Figure 3).

Discussion

This study was motivated by the somewhat counterintuitive observation in Cayetano et al. (2015) that costs and benefits of different *H. defensa* isolates to the aphid host are negatively correlated. Infection with *H. defensa* generally shortens *A. fabae*'s lifespan and thereby lifetime reproduction in the absence of parasitoids (Vorburger et al. 2013; Vorburger and Gouskov 2011), yet the magnitude of this cost is very low for the most strongly protective isolates and higher for isolates providing less protection against parasitoids (Cayetano et al. 2015). We hypothesized that costly isolates of *H. defensa* may persist in host populations because they gain some fitness from horizontal transmission, for example via parasitoid wasps (Gehrer and Vorburger 2012), aided by a high density in the host. Isolates that reach a higher titre in the host's hemolymph might increase their horizontal transmission rate as the wasp's ovipositor is more likely to be contaminated with enough bacteria to establish an infection in a new host.

Our results supported this hypothesis only partially. We confirmed that isolates differ in their titre, we demonstrated that simple stabs with symbiont-contaminated needles did indeed result in horizontal transmission, and we found that – within *H. defensa* haplotypes – higher titre was connected to increased transmission rate. Yet, our results also show that transmission rate depended strongly on *H. defensa* haplotype. Despite their low titre, isolates H76 and H101 of haplotype 1 transmitted considerably better than isolates of haplotype 2 and 3 with similarly low titres. Clearly it is not enough to enter the host in sufficient numbers – *H. defensa* also needs to overcome further hurdles on the way to establishing an infection, and the two isolates of haplotype 1 appear to be better suited to that task than others. Such among-strain variation in transmission success is also known from other bacteria, e.g. *Borrelia* in ticks (Tonetti et al. 2015), *Vibrio* in squid (Bongrand and Ruby 2019), or, potentially, *Wolbachia* in leaf cutter ants (Tolley et al. 2019).

Currently, there is no known mechanistic basis of the observed differences in *H. defensa* transmission success. It is possible that *H. defensa* isolates of haplotype 1 are better at evading the host's immune system. Given that immune activation is a costly response (Moret and Schmid-Hempel 2000; Zuk and Stoehr 2002), their limited fitness effect on the host would also be consistent with them evading recognition by the immune system. Further studies will be needed to detect how isolates of haplotype 1 interact differently with the host than isolates of haplotype 2 and 3.

The assumption that costly *H. defensa* isolates with limited benefit for the host are maintained in the population by a greater disposition for horizontal transmission must be rejected for our system. Instead, isolates of haplotype 1 seem omnipotent: They provide near-complete protection from parasitoid wasps (Cayetano et al. 2015), they hardly impair their host's offspring production and thus maximise their vertical transmission potential (Cayetano et al. 2015), and they are efficient in transmitting horizontally. This leaves us with a problem: Why do these isolates not go to fixation in the field? Likely, isolates H76 and H101 have hidden costs, costly isolates have hitherto unknown benefits, or, similar to the interaction between host and *Wolbachia,* factors such as environmental temperature or host diet influence bacterial titre (reviewed in López-Madrigal and Duarte (2019)). Both aphid and whitefly endosymbionts have been shown to influence the interaction of host and plants and to change dietary breadth (Su et al. 2015; Wagner et al. 2015). Similarly, yet unexplored interactions of *H. defensa* with host plants might lead to costs or benefits in certain habitats.

The three different aphid clones acting as recipients in our experiment did not differ in their susceptibility to infection by *H. defensa*. Two of the clones were novel hosts for the *H. defensa* isolates, while they had been associated with clone A06-407 for at least 150 host generations prior to the experiment. Despite this long lasting association – longer in fact than any that we would find in the field, where symbiont-host genotype associations are re-shuffled when *A. fabae* reproduces sexually in autumn – infection success after horizontal transmission was the same in the 'known' aphid genotype as in the two 'novel' genotypes. On one hand, this result might be influenced by the way our clonal aphid stocks are propagated. At every generation, a small number of adult aphids are used to found the next generation. The genetic drift resulting from these repeated bottlenecks may have restricted the potential for host-symbiont co-adaptation. On the other hand, this result could reflect the fact that outside of the laboratory, *A. fabae* generally reproduces by cyclical parthenogenesis (Sandrock et al. 2011). The yearly bout of sexual recombination of host genotypes may exert

selection on heritable endosymbionts to be ‘good mixers’ that can survive in any genetic background.

In addition to the transport by parasitoid wasps acting as vectors (Gehrer and Vorburger 2012), aphid endosymbionts can also be transmitted horizontally by sex (Moran and Dunbar 2006), or potentially via physical contact or via the host plant (Darby and Douglas 2003; Pons et al. 2019). The latter was a potential problem for our experiment, because 10 aphids that had been exposed to hemolymph from one donor were placed on the same leaf disc. Therefore, they fed on the same leaf during the first six days in which the infections established in the aphids. To detect whether *H. defensa* transmitted through the leaf, we placed uninfected sentinel aphids on the same leaf disc. In no case did we detect successful transmission of *H. defensa* to the sentinel aphids. It is unlikely that the sentinel clone A08-28 is resistant to infection with *H. defensa* as it was found in nature carrying a natural infection with *H. defensa*, from which it was cured in the laboratory to generate line A08-28^{H-}. It is more likely that *H. defensa* does not transmit via the leaf, or that titres of *H. defensa* in newly infected individuals were too low to allow transmission through the leaf or through physical contact. Even though we cannot exclude that plant-mediated horizontal transmission might play a role in natural settings, it did not influence the observed horizontal transmission rate in our experiment (Ewald 1987).

Interestingly, we did not observe any negative effects of newly acquired infections with *H. defensa* in terms of aphid survival directly after transfection nor in offspring production of survivors of the transfection. Generally, the survival rate in our experiment is comparable to the survival rate achieved in Niepoth et al. (2018), where *H. defensa* was transmitted via microinjection from infected pea aphid donors to pea aphid recipients that had been cured from their *H. defensa* infections. In our experiment, survival did not depend on whether hemolymph with or without *H. defensa* was transfected and did not vary significantly between different isolates of *H. defensa*. This stands in contrast to the results of Niepoth et al. (2018), where one of the two *H. defensa* isolates significantly reduced survival of all three recipients.

Previous studies have shown that infection with some *H. defensa* isolates decreases the host’s lifespan and lifetime offspring production (Cayetano et al. 2015; Niepoth et al. 2018; Vorburger and Gouskov 2011). One reason that we did not detect such a difference may be that several isolates only rarely succeeded in horizontal transmission, which decreased the

power to detect a significant difference in offspring production between recipients that did get infected by *H. defensa* and those that did not. More importantly, though, the time over which we quantified reproduction (1 week) was likely too short to detect the reproductive costs imposed by *H. defensa*, which are typically late-acting costs, mediated by curtailed lifespan rather than a reduced daily fecundity (Vorburger and Gouskov 2011). We would expect costs on offspring production to be visible with increased sample size and by assessing lifetime reproduction instead of just reproduction during the first week of adult life.

In conclusion, our experimental simulation of a known horizontal transmission route showed that both titre and haplotype of a *H. defensa* isolate determine its horizontal transmission success. This casts doubt on our initial assumption that costly isolates providing limited benefit to the host are maintained through an increased disposition for horizontal transmission, since two strongly protective and nearly avirulent isolates also showed frequent horizontal transmission. Further studies will be required to understand why such seemingly omnipotent isolates do not dominate in aphid populations.

Declarations

Acknowledgments

We thank Hugo Mathé-Hubert for help with the statistical analysis, Raffael Stegmayer and Silvan Rossbacher for help with PCR preparations and Paula Rodriguez for the long-term maintenance of the insect lines.

Author's contributions

Both authors designed the study and contributed to data analysis. HK collected the data. The first draft of the manuscript was written by HK and both authors contributed to and commented subsequent versions. Both authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Funding

This project has been funded by the Swiss National Science Foundation (Sinergia Grant No. CRSII3_154396).

Literature cited

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 67(1):1-48.

Baumann P. 1995. Genetics, Physiology, and Evolutionary Relationships of the Genus *Buchnera*: Intracellular Symbionts of Aphids. Annual Review of Microbiology. 49:55-94.

Bongrand C, Ruby EG. 2019. The impact of *Vibrio fischeri* strain variation on host colonization. Curr Opin Microbiol. 50:15-19.

Bull JJ, Molineux IJ, Rice WR. 1991. Selection of Benevolence in a Host-Parasite System. Evolution. 45(4):875-882.

Caspi-Fluger A, Inbar M, Mozes-Daube N, Katzir N, Portnoy V, Belausov E, Hunter MS, Zchori-Fein E. 2012. Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated. Proceedings Biological sciences. 279(1734):1791-1796.

Cayetano L, Rothacher L, Simon J-C, Vorburger C. 2015. Cheaper is not always worse: strongly protective isolates of a defensive symbiont are less costly to the aphid host. Proceedings of the Royal Society B. 282(1799).

Chen D-Q, Montllor CB, Purcell AH. 2000. Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid, *A. kondoi*. Entomol Exp Appl. 95(3):315-323.

Chen D-Q, Purcell AH. 1997. Occurrence and Transmission of Facultative Endosymbionts in Aphids. Current Microbiology. 34(4):220-225.

Chong RA, Moran NA. 2016. Intraspecific genetic variation in hosts affects regulation of obligate heritable symbionts. Proceedings of the National Academy of Sciences. 113(46):13114-13119.

Chrostek E, Pelz-Stelinski K, Hurst GDD, Hughes GL. 2017. Horizontal Transmission of Intracellular Insect Symbionts via Plants. Front Microbiol. 8:2237-2237.

Darby AC, Douglas AE. 2003. Elucidation of the Transmission Patterns of an Insect-Borne Bacterium. Appl Environ Microb. 69(8):4403-4407.

Ebert D. 2013. The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission. Annual Review of Ecology, Evolution, and Systematics. 44(1):623-643.

Ewald PW. 1983. Host-Parasite Relations, Vectors, and the Evolution of Disease Severity. Annual Review of Ecology and Systematics. 14:465-485.

Ewald PW. 1987. Transmission modes and evolution of the parasitismmutualism continuum. Annals of the New York Academy of Sciences. 503(1):295-306.

Ferrari J, Darby AC, Daniell TJ, Godfray HCJ, Douglas AE. 2004. Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. Ecological Entomology. 29(1):60-65.

Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. 2017. The evolution of host-symbiont dependence. Nature Communications. 8(1):15973.

Frank SA. 1992. A kin selection model for the evolution of virulence. Proceedings of the Royal Society of London Series B: Biological Sciences. 250(1329):195-197.

Fukatsu T, Nikoh N, Kawai R, Koga R. 2000. The secondary endosymbiotic bacterium of the pea aphid *Acyrthosiphon pisum* (Insecta: Homoptera). Appl Environ Microbiol. 66(7):2748-2758.

Gehrer L, Vorburger C. 2012. Parasitoids as vectors of facultative bacterial endosymbionts in aphids. Biology letters. 8(4):613-615.

Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. P Natl Acad Sci USA. 108(7):2849-2854.

Hartig F. 2020. DHAMRa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models.

Heath BD, Butcher RDJ, Whitfield WGF, Hubbard SF. 1999. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. Curr Biol. 9(6):313-316.

Henry LM, Peccoud J, Simon J-C, Hadfield JD, Maiden MJC, Ferrari J, Godfray HCJ. 2013. Horizontally transmitted symbionts and host colonization of ecological niches. Current biology : CB. 23(17):1713-1717.

Hothorn T, Bretz F, Westfall P. 2008. Simultaneous Inference in General Parametric Models. Biometrical Journal. 50:346-363.
Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software. 82(13):1-26.
Lipsitch M, Nowak MA, Ebert D, May RM. 1995. The Population Dynamics of Vertically and Horizontally Transmitted Parasites. Proceedings: Biological Sciences. 260(1359):321-327.
López-Madrigal S, Duarte EH. 2019. Titer regulation in arthropod-*Wolbachia* symbioses. Fems Microbiol Lett. 366(23).
McLean AHC, Parker BJ, Hrček J, Kavanagh JC, Wellham PAD, Godfray HCJ. 2018. Consequences of symbiont co-infections for insect host phenotypes. Journal of Animal Ecology. 87(2):478-488.
Moran NA, Dunbar HE. 2006. Sexual acquisition of beneficial symbionts in aphids. P Natl Acad Sci USA. 103(34):12803-12806.
Moran NA, Russell JA, Koga R, Fukatsu T. 2005. Evolutionary Relationships of Three New Species of Enterobacteriaceae Living as Symbionts of Aphids and Other Insects. Appl Environ Microb. 71(6):3302-3310.
Moret Y, Schmid-Hempel P. 2000. Survival for Immunity: The Price of Immune System Activation for Bumblebee Workers. Science. 290(5494):1166.
Niepoth N, Ellers J, Henry LM. 2018. Symbiont interactions with non-native hosts limit the formation of new symbioses. Bmc Evol Biol. 18(1):27.
Nowak MA, May RM. 1994. Superinfection and the evolution of parasite virulence. Proceedings of the Royal Society of London Series B: Biological Sciences. 255(1342):81-89.
Oliver KM, Russell JA, Moran NA, Hunter MS. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proceedings of the National Academy of Sciences. 100(4):1803-1807.
Peccoud J, Bonhomme J, Mahéo F, de la Huerta M, Cosson O, Simon J-C. 2014. Inheritance patterns of secondary symbionts during sexual reproduction of pea aphid biotypes. Insect science. 21(3):291-300.
Pons I, Renoz F, Noël C, Hance T. 2019. Circulation of the Cultivable Symbiont *Serratia symbiotica* in Aphids Is Mediated by Plants. Front Microbiol. 10(764).
R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
RStudio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC; Boston, MA.
Russell JA, Latorre A, Sabater-Muñoz B, Moya A, Moran NA. 2003. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. Mol Ecol. 12(4):1061-1075.
Russell JA, Moran NA. 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. Proceedings of the Royal Society B. 273(1586):603-610.
Sandrock C, Razmjou J, Vorburger C. 2011. Climate effects on life cycle variation and population genetic architecture of the black bean aphid, *Aphis fabae*. Molecular Ecology. 20(19):4165-4181.
Sandström JP, Russell JA, White JP, Moran NA. 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. Mol Ecol. 10(1):217-228.
Schmid M, Sieber R, Zimmermann Y-S, Vorburger C. 2012. Development, specificity and sublethal effects of symbiont-conferred resistance to parasitoids in aphids. Functional Ecology. 26(1):207-215.
Su Q, Oliver KM, Xie W, Wu Q, Wang S, Zhang Y. 2015. The whitefly-associated facultative symbiont *Hamiltonella defensa* suppresses induced plant defences in tomato. Functional Ecology.
Sunnucks P, Hales DF. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: *Aphididae*). Mol Biol Evol. 13:510-524.
Tolley SJA, Nonacs P, Sapountzis P. 2019. *Wolbachia* Horizontal Transmission Events in Ants: What Do We Know and What Can We Learn? Front Microbiol. 10(296).

Tonetti N, Voordouw MJ, Durand J, Monnier S, Gern L. 2015. Genetic variation in transmission success of the Lyme borreliosis pathogen *Borrelia afzelii*. Ticks and Tick-borne Diseases. 6(3):334-343.
Vorburger C, Ganesanandamoorthy P, Kwiatkowski M. 2013. Comparing constitutive and induced costs of symbiont-conferred resistance to parasitoids in aphids. Ecology and Evolution. 3(3):706-713.
Vorburger C, Gouskov A. 2011. Only helpful when required: a longevity cost of harbouring defensive symbionts. Journal of Evolutionary Biology. 24(7):1611-1617.
Vorburger C, Sandrock C, Gouskov A, Castañeda LE, Ferrari J. 2009. Genotypic variation and the role of defensive endosymbionts in an all-parthenogenetic host-parasitoid interaction. Evolution. 63(6):1439-1450.
Vorburger C, Siegrist G, Rhyner N. 2017. Faithful vertical transmission but ineffective horizontal transmission of bacterial endosymbionts during sexual reproduction of the black bean aphid, *Aphis fabae*. Ecological Entomology. 42(2):202-209.
Wagner SM, Martinez AJ, Ruan Y-M, Kim KL, Lenhart PA, Dehnel AC, Oliver KM, White JA. 2015. Facultative endosymbionts mediate dietary breadth in a polyphagous herbivore. Functional Ecology. 29(11):1402-1410.
Wickham H. 2007. Reshaping Data with the reshape Package. Journal of Statistical Software. 21(12):1-20.
Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
Wickham H, Romain F, Lionel H, Müller K. 2020. dplyr: A Grammar of Data Manipulation.
Zuk M, Stoehr Andrew M. 2002. Immune Defense and Host Life History. The American Naturalist. 160(S4):S9-S22.

Supplementary information

Supplementary Table 1 – Creation of *H. defensa*-infected A06-407 sublines. For each *H. defensa*-infected subline, the date of creation is indicated. For the clones acting as *H. defensa*-donors, collection date, site and host plant are listed. For each *H. defensa*-isolate, haplotype as determined by sequencing of the two housekeeping genes *murE* and *accD* is indicated.

A06-407 subline	Date of creation through micro-injection	Clone of origin of *H. defensa* isolate	Collection date of clone of origin	Collection site of clone of origin	Host plant of clone of origin	*H. defensa* isolate	*H. defensa* haplotype
407H9	June 2008	A06-09	08.05.2006	La Spezia, Italy	*Vicia faba*	H9	Hap2
407H15	April 2012	A06-15	08.05.2006	Ressora, Italy	*Vicia faba*	H15	Hap3
407H28	April 2012	A08-28	13.05.2008	Altstetten ZH, Switzerland	*Chenopodium album*	H28	Hap2
407H30	March 2009	A06-30	08.05.2006	Sarzana, Italy	*Vicia faba*	H30	Hap2
407H76	March 2009	A06-76	17.05.2006	La Grande Motte, France	*Chenopodium album*	H76	Hap1
407H85	September 2011	A06-85	17.05.2006	Grimaud, France	*Chenopodium album*	H85	Hap3
407H101	September 2011	A06-101	18.05.2006	Le Muy, France	*Vicia faba*	H101	Hap1
407H323	June 2008	A06-323	27.06.2006	Aesch BL, Switzerland	*Vicia faba*	H323	Hap2
407H343	September 2011	A06-343	02.07.2006	Altenhasslau, Germany	*Chenopodium album*	H343	Hap2
407H402	October 2008	A06-402	01.07.2006	St. Margret-hen SG, Switzerland	*Chenopodium album*	H402	Hap2
407HAf6	July 2008	Af6	25.05.2004	Zurich, Switzerland	*Euonymus europaeus*	HAf6	Hap2

Supplementary Table 2 – Origin of aphid clones. Collection date, site and host plant for the aphid clones free from *H. defensa*.

Aphid clone	Collection date	Collection site	Host plant	Facultative endosymbiont
A06-37	08.05.2006	Romagna, Italy	*Vicia faba*	none
A06-405	01.07.2006	St. Margrethen SG, Switzerland	*Chenopodium album*	none
A06-407	17.05.2006	St. Margrethen SG, Switzerland	*Chenopodium album*	none
A08-28^{H-}	13.05.2008	Zurich, Switzerland	*Vicia faba*	none, cured from *Hamiltonella defensa* in 2011

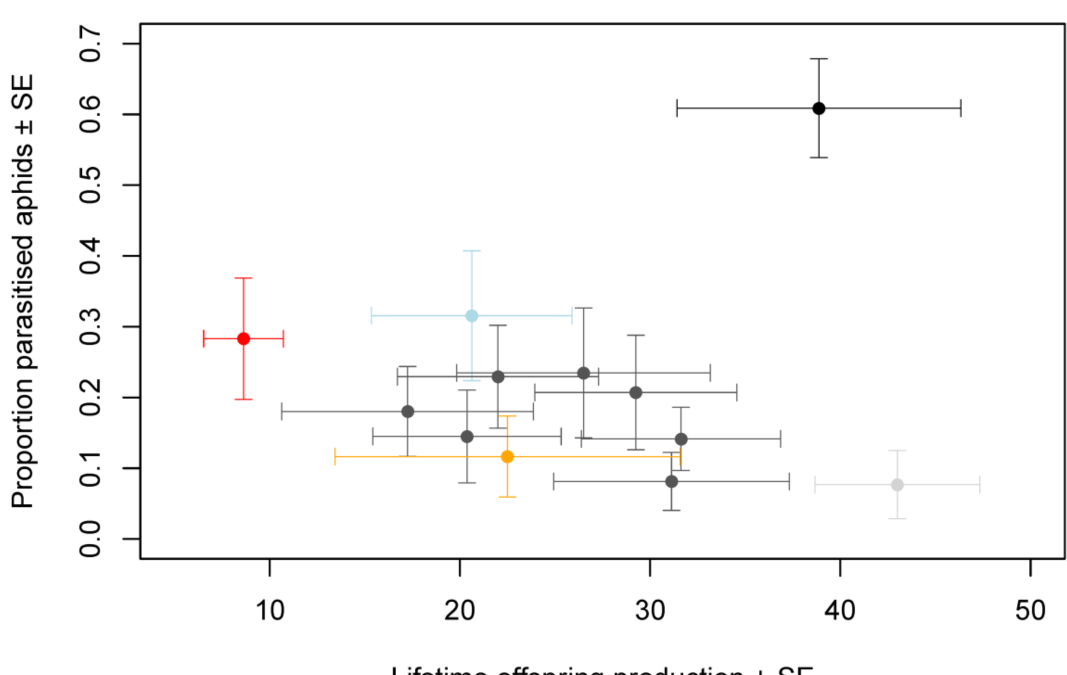
Supplementary Table 3 – Polymerase chain reaction primers and conditions. Primers and cycling conditions for PCR reactions to confirm extraction success, presence of *H. defensa* and amplification of genes used for haplotype-typing.

Primers targeting *Hamiltonella defensa*	
10F	5'-AGTTTGATCATGGCTCAGATTG-3'
T419/TO419	5'-AAATGGTATTSGCATTTATCG-3'
Primers targeting *Buchnera aphidicola*	
16SA1	5'-AGAGTTTGATCMTGGCTCAG-3'
Buch16SAfabR	5'-CTTCTTCGGGTAACGTCAAGAA-3'
Primers for haplotype typing	
murE16F	5'-ACTAACGGGAAAACCACTAATAC-3'
murE936R	5'-TTGAGAATGTCAGCGGTAATC-3'
accD291F	5'-TTCTGGAGCACAAAAAGACAC-3'
accD832R	5'-AAGGTTCAGGTTGATGAGTCAG-3'

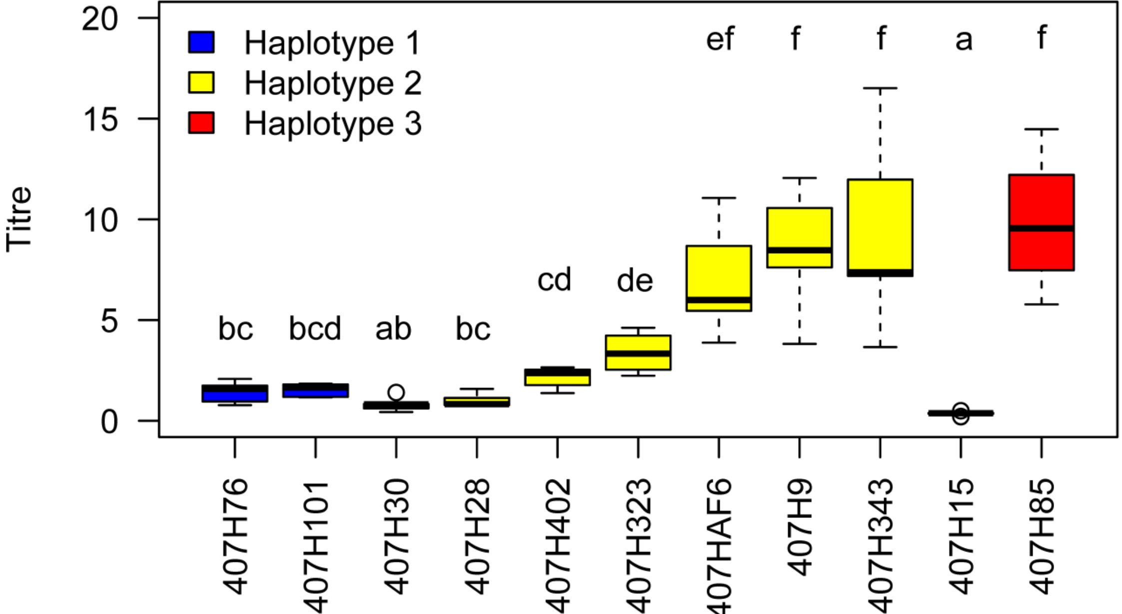
PCR program for endosymbionts			**PCR program for haplotype testing**		
Heat lid to 95°C			Heat lid to 95°C		
95°C	3 min		94°C	2 min	
95°C	30 sec	10x	94°C	30 sec	11x
65-56°C	30 sec	10x	56-46°C	50 sec	11x
72°C	60 sec	10x	72°C	50 sec	11x
95°C	30 sec	25x	94°C	30 sec	25x
55°C	30 sec	25x	45°C	50 sec	25x
72°C	60 sec	25x	72°C	60 sec	25x
72°C	6 min		72°C	6 min	
10°C	∞		4°C	∞	

Supplementary Table 4 – Results of a generalised linear mixed effects model for the transmission rate of different *H. defensa* isolates to different recipients (37H0, 405H0, 407H0). Model predictors were recipient, average titre that an isolate reaches in the donor aphid and haplotype of the isolate (haplotypes 1, 2 and 3). The aphid subline acting as donor during horizontal transmission ('donor') and experimental batch were treated as a random effect. The model did not contain data from batch 1.

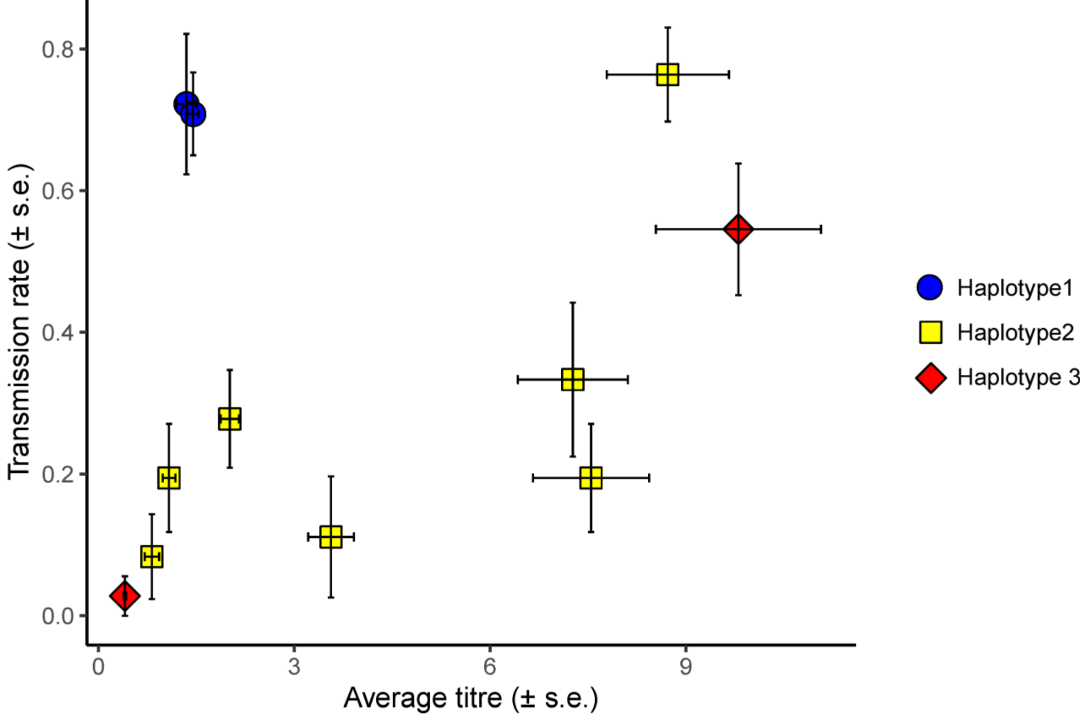
	Effect	LR χ2	df	p-value
Random:	donor	18.76	1	<0.001
	batch	2.54	1	0.111
Fixed:	titre of H. defensa isolate	8.34	1	0.004
	haplotype of H. defensa isolate	11.48	2	0.003
	recipient clone	1.87	2	0.393
	titre : recipient	10.88	2	0.004
	recipient : haplotype	10.38	4	0.034



Supplementary Figure 1 – Figure adapted from Cayetano et al. (2015): Impact of infection with different *H. defensa* isolates on offspring production (cost) and parasitation success of the parasitoid wasp *Lysiphlebus fabarum* (benefit) in aphid clone A06-407. The aphid clone is uninfected (H0) or infected with different *H. defensa* isolates). For clarity, some isolates are marked by different colours: The highly protective and avirulent isolate H76 (light grey) of haplotype 1, one isolate of haplotype 2 (H402, orange), and the two isolates of haplotype 3 (H15 in light blue, and the very costly H85 in red). Error bars depict the standard error.



Supplementary Figure 2 – Titre of *H. defensa*, expressed as the ratio of gene counts from the two housekeeping genes *dnaK* (*H. defensa*) and *EF1α* (*A. fabae*), was measured from pools of three aphids for each of the five batches processed in the experiment using qPCR. The *H. defensa* isolates are coloured according to haplotype (blue = haplotype 1, yellow = haplotype 2, red = haplotype 3). Different letters indicate significant differences between titres.



Supplementary Figure 3 – Average titre of different *H. defensa* isolates in donor aphids plotted against the average transmission rate of the isolate without data from aphids of batch 1. Transmission rate corresponds to the number of recipients out of three or four in which a *H. defensa* isolate successfully established after horizontal transmission, i.e. was propagated to the recipient's offspring. In this figure, transmission rate is averaged over three different recipients (37H0, 405H0 and 407H0) and four experimental batches (batch 1 was removed from the analysis). Error bars indicate the standard error and the combination of colour and symbols indicates the haplotype of the *H. defensa* isolate (blue circle = haplotype 1, yellow square = haplotype 2, red diamond = haplotype 3).

Chapter 3

Similar cost of *Hamiltonella defensa* in experimental and natural aphid-endosymbiont associations

Heidi Kaech[1,2], Stephanie Jud[2], Christoph Vorburger[1,2]

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[2]D-USYS, Department of Environmental Systems Science, ETH Zürich, Switzerland

Publication in preparation

Abstract

Endosymbiont-conferred resistance to parasitoids is common in aphids but comes at a cost to the host in the absence of parasitoids. In black bean aphids (*Aphis fabae*), this was demonstrated by introducing eleven isolates of the protective symbiont *Hamiltonella defensa* into previously uninfected clones, which led to reduced lifespan and lifetime reproduction. However, it has been suggested that this approach could overestimate the realised costs of the endosymbiont, because transfection creates new and potentially maladapted host-symbiont combinations that would be eliminated by natural selection in the field. Here we show that removing these *H. defensa* isolates from their natural host clones results in a fitness gain that is comparable to the fitness loss from their introduction into new clones, thus validating the transfection approach. From failures to cure two naturally infected clones we discovered that one haplotype of *H. defensa* appears to be resistant to the antibiotic cefotaxime, to which other haplotypes are susceptible. By comparing our fitness estimates of transfected aphid lines with estimates reported in previous publications using the same lines, we could also show that symbiont-induced costs can fluctuate over time. Thus, costs estimated after extended culture in the laboratory may not always be representative of the costs at the time of collection in the field. **Keywords:** *Aphis fabae*, cost, common genetic background, *Hamiltonella defensa*, fitness, symbiont

Introduction

The survival of aphids depends on the bacterial symbionts they are associated with. Curing aphids from their primary endosymbiont, the γ-proteobacterium *Buchnera aphidicola,* has a fatal impact. Without *B. aphidicola* providing them with essential amino acids, aphids grow slowly and fail to reproduce (Douglas 1998; Hansen and Moran 2011; Sasaki et al. 1991). Association with other, 'secondary' bacterial endosymbionts is not as vital for aphids, but can substantially increase their survival under specific ecological conditions. Secondary endosymbionts can, for example, protect their hosts against natural enemies (Ferrari et al. 2004; McLean et al. 2020; Oliver et al. 2003; Vorburger et al. 2009) or reduce the negative impact of heat stress (Chen et al. 2000; Russell and Moran 2006). Despite these benefits, secondary symbionts only occur at intermediate frequencies in aphid populations (Smith et al. 2015; Vorburger and Rouchet 2016). Their spread seems to be constrained by costs; but these can be difficult to detect. Some assays of fitness components reported no systematic costs of infection with secondary endosymbionts (Łukasik et al. 2013b; Russell and Moran 2006), while in others such costs were clearly visible (Cayetano et al. 2015; Chen et al. 2000; Leybourne et al. 2020; Vorburger et al. 2013; Vorburger and Gouskov 2011). Additionally, costs can seem absent in components of fitness assays but become apparent when aphids with and without secondary endosymbionts compete for resources in population cages (Dykstra et al. 2014; Oliver et al. 2008; Rossbacher and Vorburger 2020). In a meta-analysis of 57 papers, Zytynska et al. (2019) provided an overview over the sleuth of studies on the topic. They found that the protection against parasitism comes – across all analysed aphid species and secondary symbionts – with costs such as increased age at reproduction, shorter lifespan and decreased fecundity. However, they also found substantial variation between experiments that compare naturally infected to naturally uninfected aphids, and experiments that use experimentally manipulated aphid lines, that were cured from or transfected with an endosymbiont.

Experiments that compare naturally infected to naturally uninfected aphids are rarer and find more variable but also less severe costs of secondary symbionts (Zytynska et al. 2019). Either, the natural environment selects for aphid-symbiont-associations with low costs, or this is a result of low sample size and of confounding several sources of variation: the inherent fitness of different aphid clones (visible in the variation of fitness among naturally uninfected aphids in Castañeda et al. 2010; Ferrari et al. 2007; Vorburger et al. 2009), the inherent cost of different endosymbiont isolates (Cayetano et al. 2015; Łukasik et al. 2013a), the genotype-by-

genotype interaction between aphid clone and endosymbiont isolate (Ferrari et al. 2007; Parker et al. 2017; Vorburger and Gouskov 2011), and genotype-by-environment interactions (e.g. host plant effects: Ferrari et al. 2007; McLean et al. 2011). While the transfer of isolates of secondary endosymbionts into a common genetic background helps disentangle these different sources of variation, the approach may result in a biased perception of the costs associated with the infection. Experimental transfections create new combinations of secondary endosymbionts and aphid genotypes similar to when aphids reproduce sexually. But unlike in nature, these new combinations created in the laboratory are not tested and optimised by natural selection. As a result, unfavourable combinations of host and symbiont genotypes, which would quickly disappear in nature, might be used to assess the costs of secondary endosymbionts. Experiments with experimental host-symbiont combinations may thus overestimate the realised cost of secondary endosymbionts in nature. This can be overcome by curing aphid clones of their natural infections with antibiotics, but care must be taken to ensure that there are no lingering negative effects of the antibiotic treatment on the host or its primary endosymbiont *B. aphidicola.* Otherwise there is a risk of underestimating the cost of secondary symbionts, as the cured aphid suffers from reduced fitness due to the antibiotic-treatment.

In this work, we addressed these two issues using a well-characterised set of 11 isolates of the secondary endosymbiont *Hamiltonella defensa* of the black bean aphid, *Aphis fabae* (see Cayetano et al. 2015; Vorburger and Gouskov 2011). When these isolates are transfected to two naturally uninfected aphid clones, i.e. into common genetic backgrounds, they provided provides protection from parasitoid wasps to *A. fabae* (Cayetano et al. 2015; Vorburger et al. 2009) but also cause costs to lifespan and offspring production (Cayetano et al. 2015; Vorburger et al. 2013; Vorburger and Gouskov 2011). In this experiment, we compared the cost of the 11 *H. defensa* isolates in the two naturally uninfected host genotypes, which they had been transfected into, to the cost in their naturally associated host genotypes. We also tested whether antibiotic treatment *per se* impacts the fitness of aphid clones by curing naturally infected aphid clones and reinfecting them with their own isolate of *H. defensa*. The two approaches allowed us to assess how well experiments in a common genetic background reflect costs of different *H. defensa* isolates in natural associations found in the field.

Methods

Aphid clones and *H. defensa* isolates

We measured lifetime offspring production and age at death for 51 lines of *A. fabae* (Supplementary Table 1). There were 12 sublines each of clones A06-405 and A06-407. These two clones were originally free from secondary endosymbionts (subline 407H0, which corresponds to the treatment 'uninfected A06-407', and subline 405H0, the 'uninfected A06-405') and had been microinjected with 11 different *H. defensa* isolates (H9 to HAf6) from other *A. fabae* clones between 2008 and 2012 to form sublines 407H15 to 407HAf6 ('infected A06-407') and 405H15 to 405HAf6 ('infected A06-405') (Supplementary Table 1). There were also the 11 aphid clones that the different *H. defensa* isolates were associated with in nature (A06-09 to Af6, 'naturally infected'), as well as 10 cured (A06-09^{H-} to Af6^{H-}, 'cured') and six reinfected sublines (A06-09$^{H.reinf}$ to AF6$^{H.reinf}$, 'reinfected') (Supplementary Table 1). Since collection or creation, all clones and sublines have been maintained in asexually reproducing colonies on broad beans (*Vicia fabae*) at 18-20°C and a 16 h photoperiod.

Based on partial sequences of two bacterial housekeeping genes, *murE* and *accD,* the different *H. defensa* isolates in this experiment can be grouped into three haplotypes (Supplementary Table 1): Haplotype 1 comprises H76 and H101, haplotype 2 comprises H9, H28, H30, H323, H343, H402 and AF6, and haplotype 3 comprises H15 and H85 (Cayetano et al. 2015). The division into these three haplotypes has been confirmed by sequencing additional genes (Youn Henry, personal communication).

Curing aphid clones from *H. defensa*

Most aphid clones were cured from their *H. defensa* infection by oral uptake of the antibiotic cefotaxime (LGC Standards). Broad bean leaves were inserted through a hole in the lid into 1.5 ml Eppendorf tubes containing a solution of 1 mg/ml cefotaxime in tap water. Of each naturally infected aphid clone, six 3- to 4-day-old aphids were placed on the leaf. The Eppendorf tube was encased in a Falcon tube, which was sealed with a foam plug. Each clone was treated in two sequential batches, respectively, with three experimental blocks in the first and one or two experimental blocks in the second batch. Aphids fed on the antibiotic-laced plant sap for 48 h in the first batch and for 72 h in the second batch before being placed individually on *V. faba* seedlings at 18°C to reproduce. Twenty-nine days after exposure to antibiotics, three adult daughters of the aphids that survived the antibiotic treatment were allowed to reproduce overnight on *V. faba* seedlings before their DNA was extracted and

tested for presence of *B. aphidicola* and *H. defensa.* Two aphid clones, A06-76 and A06-101, could not be cured with the protocol described above. Those we then tried to cure in a less systematic manner. We tested different dosages of meropenem (Adooq Bioscience), phosphomycin disodium (Sigma) and a mixture of cefotaxime (Dr. Ehrenstorfer GmbH), gentamycin sulfate (PanReac AppliChem) and ampicillin (Calbiochem) (McLean et al. 2011), applied either orally or through microinjection. Eventually, one *H. defensa*-free aphid of clone A06-101 was obtained. Its mother had been feeding for three days on a mixture of antibiotics (100 µg/ml ampicillin, 50 µg/ml cefotaxime and 50 µg/ml gentamycin, dissolved in tap water). Despite extensive trials, we did not manage to cure clone A06-76 from its *H. defensa* infection.

Reinfection of aphid clones

Approximately 12 generations after antibiotic curing, we tried to reinfect all cured sublines with the *H. defensa* isolate that they originally were associated with. Four- to five-day-old aphids were injected with hemolymph from adult donors under CO_2-anaesthesia using a FemtoJet 4i microinjector and placed in insect breeding dishes (Ø 5 cm), which contained a broad bean leaf disc (Ø 4 cm) on 1% agar. The aphids were maintained at 21°C and a 16 h photoperiod until they died. Their last three offspring were allowed to reproduce on broad bean seedlings before being tested for presence of *H. defensa.* Reinfection was successful for six clones (Supplementary Table 2). These were used to compare the fitness of naturally infected aphids with the fitness of cured and reinfected aphids of the same host-symbiont combination.

Experimental procedures

Approximately 16 generations after the initial antibiotic treatments and four generations after reinfections, we estimated lifespan and lifetime reproduction of all 51 aphid lines described in Supplementary Table 1. To prevent carry-over of environmental maternal effects from stock cultures, experimental setup was as follows: Two generations before the start of the experiment, 51 bean seedlings – one for each line – were infested with 5 adult aphids each. The adults were allowed to produce offspring for two days before being removed singly into Eppendorf tubes and frozen at -20°C until DNA extraction (see Molecular methods below). When the offspring had reached adulthood, they were used to set up ten experimental blocks. Note that the ten experimental blocks were set up in four batches (3+3+2+2) on four consecutive days. For each block, we took a potted 1-week-old broad bean seedling per aphid

line and infested it with three adult aphids. The seedlings were covered with a cellophane bag, which was secured to the pot with a rubber band, and placed on a tray in random order (randomized complete blocks). Thus, each of the ten blocks consisted of one tray containing one replicate of all 51 aphid lines. The trays were placed in a climatized room at 21°C and a 16 h photoperiod. The adults were allowed to reproduce overnight before being removed and discarded. At that point, the experimenter was blinded to the line identity of the replicate colonies. The experimental generation was started after nine days, when the offspring had reached adulthood: Five reproducing adults per replicate were transferred to new bean seedlings. After six hours the adults were discarded and all but one offspring were removed from the plant. Five days after setup, aphids were checked for survival and transferred to new plants. Old plants were discarded. From this point onwards, survival was assessed every second day and the number of offspring produced was counted every fifth day when the aphids were transferred to new plants. All aphids were followed to the end of their life. Thirteen aphids that were killed or lost during transfers were removed from the analysis.

Molecular methods

DNA was extracted from the 51x5 aphids collected two generations before the start of the experiment using a 'salting out' protocol (Sunnucks and Hales 1996). Extraction success was verified by amplifying part of the 16S rRNA gene of *B. aphidicola*, the obligate endosymbiont present in all aphids, using specific primers. The presence/absence of *H. defensa* was also determined by diagnostic PCR with specific primers for the same gene. Primers and cycling conditions are detailed in Supplementary Table 2. Amplicons were run and visualized by capillary electrophoresis on a QIAxcel Advanced System (Qiagen AG, Hombrechtikon, Switzerland). Aphids were genotyped at eight microsatellite loci (Coeur d'Acier et al. 2004; Sandrock et al. 2011) and allele scoring was done with GeneMarker v2.4.0. One of the five individual aphids of each line was selected at random to identify *H. defensa* haplotypes through amplification of *murE* and *accD* gene fragments (primers and cycling conditions in Supplementary Table 2) and Sanger sequencing of the amplicon by a commercial provider (GATC Biotech AG, Köln, Germany). We found that due to an experimental error, all replicates of the aphid subline 405H9 were actually subline 407H0. These replicates were reassigned to subline 407H0 for data analysis.

Comparison to previous experiments

Our lab had assessed the impact of different *H. defensa* isolates on lifespan and reproduction of the two aphid clones A06-405 and A06-407 on previous occasions. We compared the results reported from these prior experiment in Vorburger and Gouskov (2011) and Cayetano et al. (2015) to our data. All three experiments were conducted in complete random block designs with aphids using broad bean seedlings in 0.07 l plastic pots as host plants, but each experiment was conducted in a different climate chamber and by different experimenters. Life history traits were assessed at a temperature of 21°C in this study, and at 20°C in Cayetano et al. (2015) and Vorburger and Gouskov (2011). The photoperiod was always set to 16 h and aphids were transferred to new host plants every fourth or fifth day. In 2018 (chapter 2), we measured the titre of the different *H. defensa* isolates in aphids of clone A06-407 raised at 18°C. Titre was defined as the ratio of *H. defensa dnaK* to *A. fabae EF1α* copy numbers measured by qPCR.

Statistical analysis

Statistical analyses were done in RStudio v1.1.463 (RStudio Team 2020) and R v3.5.1 (R Core Team 2018) using the packages survival v3.1-12 for survival plots (Therneau 2020b), coxme v2.2-16 (Therneau 2020a) for Cox mixed-effect models, permuco v1.1.0 for permutation of factorial ANOVAs (Frossard and Renaud 2019), ggplot2 v3.3.2 (Wickham 2016) and gridExtra v2.3 (Auguie 2015) for producing figures, and reshape2 v0.8.8 (Wickham 2007) and plyr v1.8.6 (Wickham et al. 2020) for data wrangling.

Survival data were analysed with a Cox mixed-effect model testing for the effect of treatment (cured, reinfected, naturally infected aphid lines as well as experimentally infected (transfected) and naturally uninfected A06-405 and A06-407 clones) with experimental block, clone and *H. defensa* isolate as random effects.

To calculate the loss of lifetime and offspring caused by *H. defensa*, we subtracted the lifespan of each replicate individual of infected clones from the average lifespan of all replicates of its uninfected counterpart (i.e. for each individual of clone A06-405 infected with *H. defensa* isolate H101 we subtracted its lifespan from the average lifespan of all uninfected A06-405). Note that the isolate H76 was excluded from these analyses, as its naturally associated aphid clone A06-76 could not be cured from *H. defensa*. Since the residuals of linear mixed models deviated significantly from uniformity, we used permutation ANOVAs with the "dekker" method, which is more appropriate for unbalanced designs, and 100'000

permutations to analyse influence of genetic background (natural host clone or the two experimentally infected host clones), *H. defensa* isolate, the interaction between genetic background and isolate, as well as experimental block on the number of offspring and lifespan lost through infection. In two additional permutation ANOVAs, we compared lifespan and lifetime reproduction of six aphid lines for which reinfection was successful between reinfected and naturally infected lines. Treatment (reinfected or naturally infected), isolate, the interaction between treatment and isolate as well as experimental block were treated as fixed effects.

Results

Antibiotic treatment: A cefotaxime-resistant haplotype of *H. defensa*

Of 282 antibiotic-treated aphids whose mothers were fed with cefotaxime-laced plant sap, 177 (62.77%) were successfully cured of *H. defensa*. The success of antibiotic treatment was significantly different among *H. defensa* haplotypes (χ^2=120.44, df=2, p<0.001). The percentage of cured individuals was high and very similar for aphids infected with *H. defensa* of haplotype 2 (78.9%) and haplotype 3 (78.2%) (χ^2=0, df=1, p=1), whereas not a single individual of clones A06-76 and A06-101 lost the infection with *H. defensa*. These clones carry *H. defensa* haplotype 1, which appears to be resistant to treatment with cefotaxime. Applying a brute force approach with multiple antibiotics at different concentrations (see Methods) we later obtained one individual of clone A06-101 that lost its *H. defensa* infection, but we did not manage to cure A06-76.

Costs of infection with *H. defensa*

Of 510 aphids followed over the course of their life, 13 (2.5%) were lost due to accidents and removed from the analysis. All aphids survived to at least the first assessment of survival at five days of age, median survival was 23 days and the last aphid died at 53 days of age. Aphids produced most of their offspring in the first 20 days of their lives, reaching on average 23.6 ± 0.8% (SE) of their maximal offspring production within ten days, 71.0 ± 0.9% within 15 days and 93.4 ± 0.4% within 20 days.

Different aphid lines varied in their lifespan. A Cox mixed effects model revealed a highly significant random effect of clone (χ^2=15.85, df=1, p<0.001) and *H. defensa* isolate (χ^2=95.83, df=1, p<0.001), while the random effect of the experimental block was not significant (χ^2=0.05, df=1, p=0.817). An overview over lifespan and offspring production of all lines in the experiment is provided in Supplementary Figure 1 & 2, respectively. Generally, infection

with *H. defensa* was costly (Table 1, Figure 1 A). Contrasts comparing different treatments showed that aphids infected by *H. defensa* died significantly earlier than uninfected aphids (Table 2). There was, however, no significant change in survival due to reinfection or between cured and naturally uninfected aphids (Table 2).

Table 1 – Maximal, median and mean longevity and reproduction in aphids that were infected with their naturally associated *H. defensa* isolate, cured from their infection, or cured and reinfected with their associated *H. defensa,* as well as longevity and reproduction in naturally uninfected aphids into which different *H. defensa* isolates were experimentally infected (transfected) or that were tested in their natural uninfected state. Treatments are sorted by infection status, with infected aphids on the top. The number of replicates in each treatment is indicated.

Treatment	Replicates	Age at death (in days)			Number of offspring		
		Maximum	Median	Mean	Maximum	Median	Mean
reinfected	57	39.0	23.0	22.8	99.0	77.0	72.0
naturally infected	108	43.0	23.0	22.7	105.0	75.5	65.7
transfected A06-405	100	39.0	19.0	19.7	94.0	59.5	54.3
transfected A06-407	105	37.0	15.0	18.3	94.0	46.0	48.0
cured	97	53.0	43.0	39.6	106.0	88.0	82.9
naturally uninfected A06-405	10	45.0	39.0	35.0	86.0	80.5	73.1
naturally uninfected A06-407	20	49.0	39.0	34.8	96.0	68.0	60.9

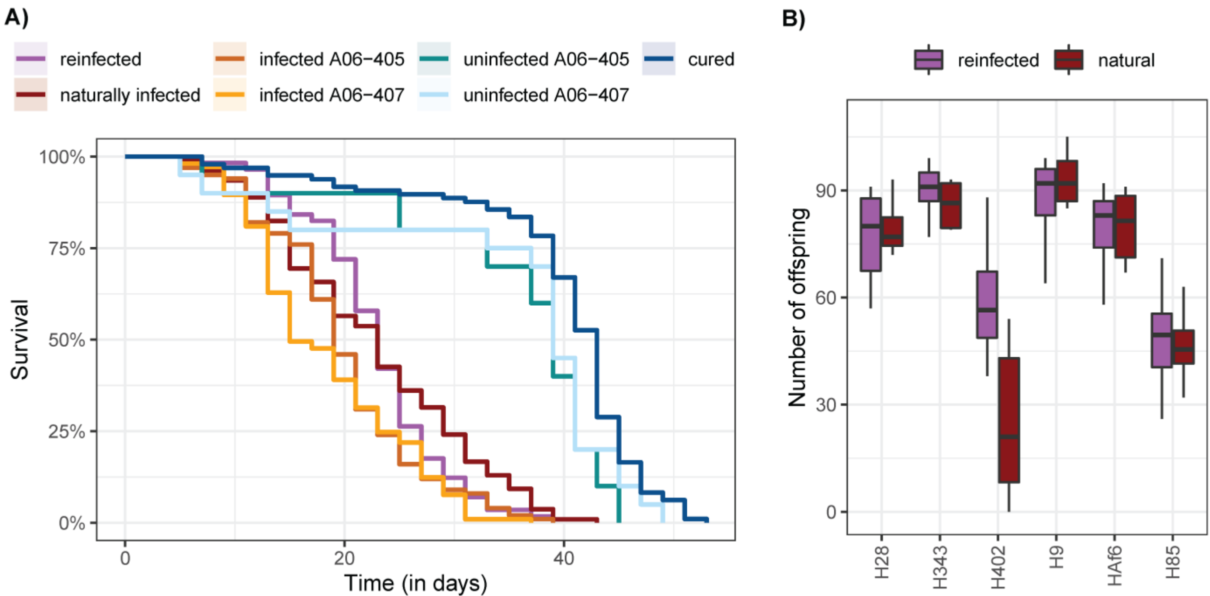


Figure 1 – (A) Survival of aphids naturally infected by *H. defensa* (dark red), cured from *H. defensa* (dark blue), cured and reinfected (magenta), naturally uninfected (uninfected A06-405, cyan; uninfected A06-407, light blue) and transfected with different *H. defensa* isolates (infected A06-405, dark orange; infected A06-407, orange). (B) Reproduction of aphids naturally infected by *H. defensa* (dark red) and aphids cured from and reinfected with *H. defensa* (magenta).

Averaged over 10 isolates (H76 was excluded because no cured line could be obtained), aphids infected with *H. defensa* lost 16.84 offspring and their lifespan was shortened by 16.59 days compared to uninfected aphids. For a comparison between isolates see Table 3.

Table 2 – Cox regression on the influence of treatment on lifespan was followed by post hoc tests with comparisons specified as custom contrasts between different experimental treatments (reinfected, naturally infected and cured aphid clones as well as uninfected and experimentally infected (transfected) A06-405 and A06-407). Resulting p-values are considered significant if they are below a p-value of 0.0125, which corresponds to a Bonferroni correction.

Post hoc comparison	Estimate	Std. error	z value	p-value
Infected vs. uninfected lines	-3.145	0.527	-5.971	**<0.001**
Reinfected vs. naturally infected lines	-0.086	0.183	-0.473	0.636
Cured vs. naturally uninfected lines	0.728	1.076	0.676	0.499
Naturally infected vs. transfected lines	0.804	0.456	1.763	0.078

Table 3 – Impact of different *H. defensa* isolates on the lifespan and reproduction of the aphid clone they were naturally associated with, and on naturally uninfected aphid clones that they were transfected into (A06-405 and A06-407). Costs are averaged over all replicates and expressed in days of life and number of offspring lost (or gained, if there is a negative value) due to infection with *H. defensa.* Missing data, caused by failure to include line 405H9 into the experiment, are indicated by 'NA'. Aphids infected with isolate H76 are not included due to failure to cure the A06-76 subline.

H. defensa isolate	Average cost on lifespan (in days)			Average cost on reproduction		
	naturally associated clone	clone A06-405	clone A06-407	naturally associated clone	clone A06-405	clone A06-407
H101	17.2	6.6	13.6	28.1	5.3	4.0
H9	15.6	NA	22.0	8.5	NA	35.3
H28	12.4	17.8	15. 3	0.6	20.1	-4.1
H30	7.4	13.2	14.6	-4.8	1.4	-4.6
H323	27.1	18.8	22.6	52.6	22.9	34.5
H343	15.8	14.0	20.2	0.9	-0.9	23.1
H402	28.2	17.2	12.4	60.6	18.2	-3.4
HAf6	11.6	14.6	21.2	-0.7	36.7	28.2
H15	8.2	7.6	8.2	4.7	-5.9	-14.57
H85	30.6	23.6	25.0	39.3	56.2	48.6

The amount of lifespan lost depended significantly on the *H. defensa* isolate ($F_{9,247}$=29.14, $p_{permutation}$<0.001) but not on whether the isolate was associated with its natural aphid genetic background or experimentally transferred to one of the originally uninfected clones ($F_{2,247}$=0.01, $p_{permutation}$=0.869). However, there was a significant interaction between genetic background and *H. defensa* isolate ($F_{18,247}$=6.53, $p_{permutation}$<0.001), indicating that the impact of different *H. defensa* isolates varied significantly depending on which aphid clones they were associated with. Lifespan did not vary between experimental blocks ($F_{9,247}$=0.96, $p_{permutation}$=0.475). Similarly, the number of offspring lost depended significantly on *H. defensa* isolate ($F_{9,247}$=21.41, $p_{permutation}$<0.001) and the interaction between isolate and background ($F_{18,247}$=7.00, $p_{permutation}$<0.001), but there was no significant main effect of genetic background ($F_{2,247}$=1.07, $p_{permutation}$=0.146) or experimental block ($F_{9,247}$=0.97, $p_{permutation}$=0.464).

Natural infections vs. reinfections: no negative effect of antibiotic treatment

A separate analysis of six clones, for which naturally infected as well as cured and reinfected lines were available, provided no evidence that the antibiotic curing has any long-lasting negative effects on aphid fitness. The clones varied significantly in average lifetime reproduction and lifespan, and there was a significant or near-significant overall difference between naturally infected and reinfected lines (Table 4). There was also a significant interaction between aphid clone and treatment (natural or reinfected) for reproduction and lifespan (Table 4). The significant main effect of treatment and the significant interaction between treatment and isolate were driven exclusively by clone A06-402 (Figure 1 B). Its naturally infected line exhibited conspicuously low fitness, which was improved in the cured and reinfected line. When clone A06-402 was excluded from the analysis, the among-clone variation remained significant, but there was no significant effect of treatment (Table 4) and no clone-by-treatment interaction (Table 4).

Table 4 – Results of permutation ANOVAs assessing the influence of antibiotic cure with subsequent reinfection on aphid fitness. Response variables were (a) lifespan or (b) lifetime reproduction. Predictors were treatment (natural or reinfected), *H. defensa* isolate and experimental block. The models were run with or without clone A06-402. P-values are based on the permutation method "dekker" with 100'000 permutations

	Source	All clones			Clone A06-402 excluded		
		df	F	P_{perm}	df	F	P_{perm}
(a)	Lifespan						
	Experimental block	9	0.57	0.835	9	0.49	0.901
	Treatment (nat. vs. reinfected)	1	3.77	0.054	1	0.01	0.910
	Isolate	5	26.65	**<0.001**	4	30.74	**<0.001**
	Treatment × isolate	5	4.91	**<0.001**	4	0.70	0.609
(b)	Lifetime reproduction						
	Experimental block	9	1.22	0.285	9	1.30	0.236
	Treatment (nat. vs. reinfected.)	1	5.81	**0.016**	1	0.39	0.544
	Isolate	5	28.21	**<0.001**	4	20.21	**<0.001**
	Treatment × isolate	5	3.49	**0.005**	4	0.50	0.752

Costs of *H. defensa* changed over time

The *H. defensa*-infected and uninfected sublines of aphid clones A06-405 and A06-407 have been maintained clonally in our lab for approximately a decade. Their life history traits have previously been assessed by Cayetano et al. (2015) and Vorburger and Gouskov (2011). This provided an opportunity to examine the consistency of *H. defensa*-induced costs over many years of laboratory culture. Generally, lifespan (Figure 2) and lifetime reproduction (Supplementary Figure 3) correlate positively across studies, with the two experiments conducted at 20°C (2015 and 2011) correlating more closely with each other than with the data from this experiment, which was conducted at 21°C. Despite these overall correlations,

costs of different *H. defensa* isolates on their host's reproduction (Figure 3 A, C, E) and lifespan (Figure 3 B, D, F) varied considerably over time. For example, costs of isolate H30 on both lifespan and reproduction decreased successively in A06-407, while costs of H323 on reproduction increased over time.

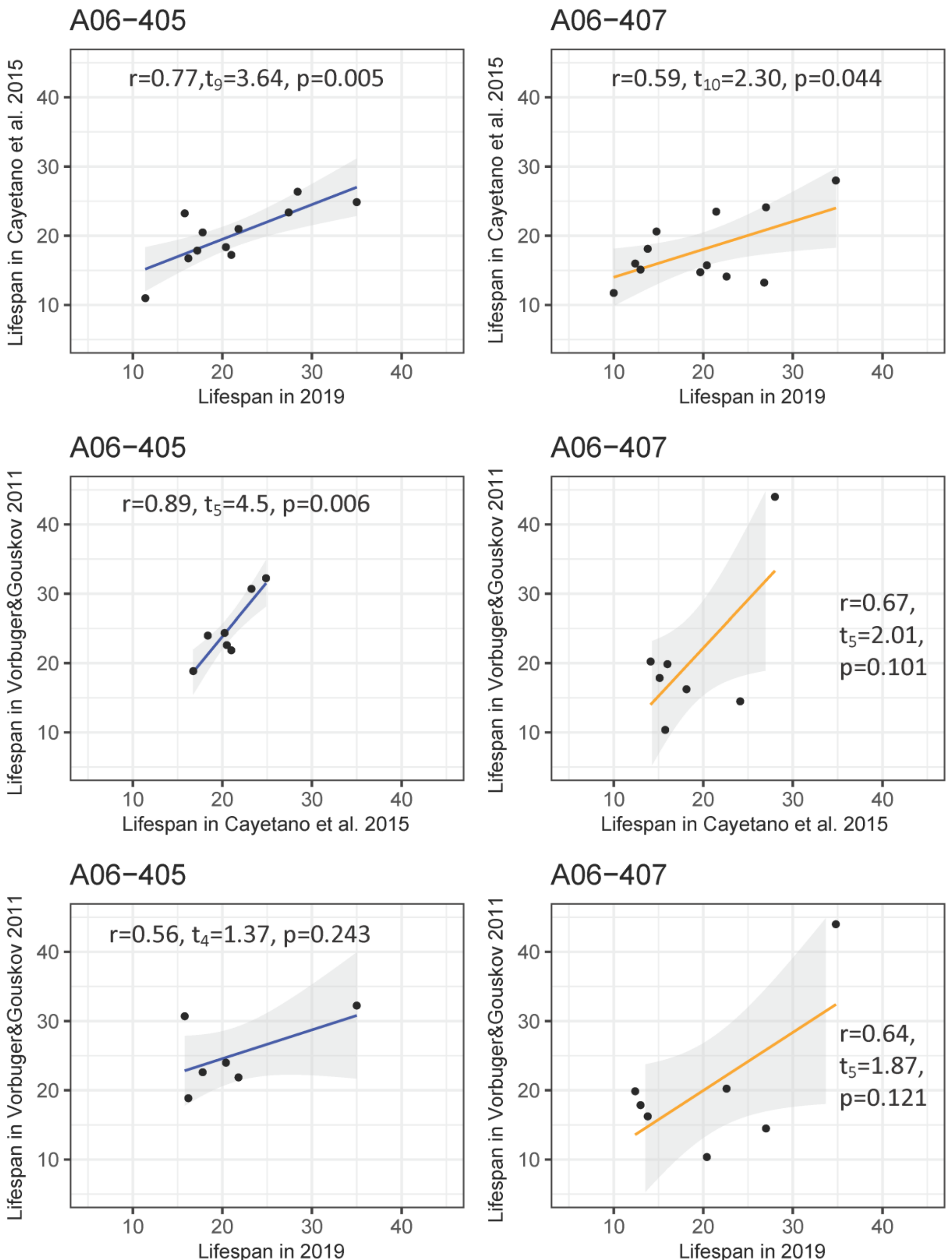


Figure 2 – Pearson's product-moment correlation of lifespan of aphid clones A06-405 and A06-407 with or without *H. defensa*-infection among three different experiments (Vorburger and Gouskov (2011), Cayetano et al. (2015) and this study).

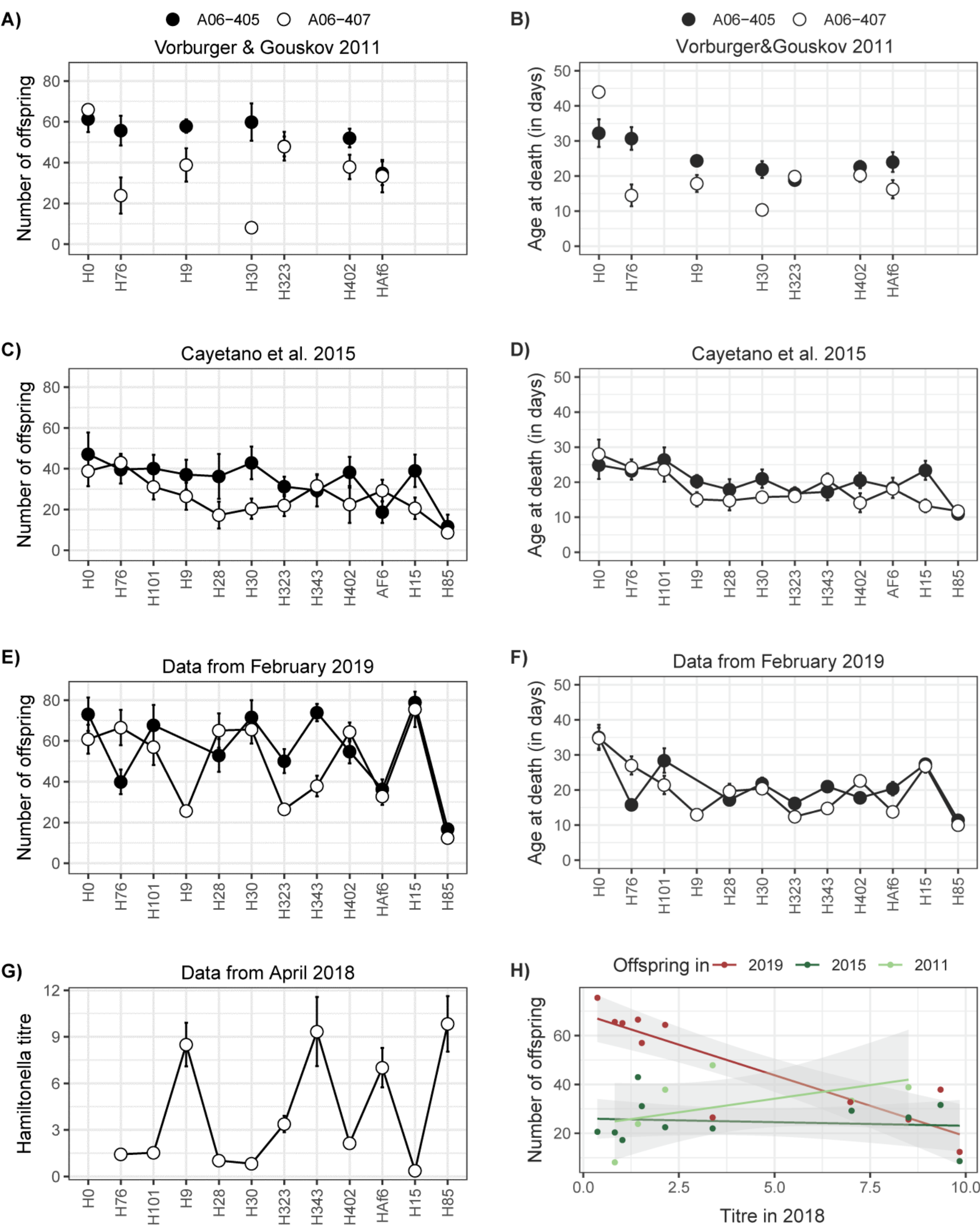


Figure 3 – Costs of different *H. defensa* isolates varied between 2011 and 2019. The two aphid clones A06-405 (black circles) and A06-407 (white circles) have been maintained in clonal lines either uninfected with *H. defensa* (H0) or infected with different *H. defensa* isolates (H16 to H85). Lifetime reproduction of the lines was measured in (A) 2011, (C) 2015 and (E) 2019, and lifespan was measured in (B) 2011, (D) 2015 and (F) 2019. (G) Titre defined as the ratio of *H. defensa dnaK* to *A. fabae EF1α* copy numbers was measured for different *H. defensa* isolates in clone A06-407 in 2018. (H) Titre of different *H. defensa* isolates measured in 2018 in A06-407 compared to reproduction of *H. defensa* –infected A06-407 in 2011, 2015 and 2019. Error bars indicate standard errors.

In 2018, we also measured the titre of different *H. defensa* isolates in clone A06-407 (Figure 3 G and Chapter 2). The *H. defensa* titres estimated in 2018 show a strong negative correlation with the number of offspring produced by *H. defensa*-infected A06-407 in 2019 (Pearson's product-moment correlation: $r=-0.87$, $t_9=-5.35$, $p<0.001$) and with lifespan in 2019 ($r=-0.82$, $t_9=-4.28$, $p=0.002$). But titre does not correlate with fitness components measured in 2015 and 2011 (Figure 3 G), neither for reproduction (2015: $r=-0.12$, $t_9=-0.37$, $p=0.717$, 2011: $r=0.50$, $t_4=1.17$, $p=0.308$) nor for lifespan (2015: $r=-0.13$, $t_9=-0.40$, $p=0.696$, 2011: $r=0.34$, $t_4=0.77$, $p=0.486$).

Discussion

There is concern that estimating costs of infection with the protective endosymbiont *H. defensa* from experimentally infected clones generates an upwardly biased perception of these costs in natural populations (Poulin and Keeney 2008; Vorburger 2014). On one hand, costs estimated from arbitrary combinations of host and symbiont genotypes are representative of the range of potential negative effects of *H. defensa*. On the other hand, they might differ significantly from the realized costs, as host-symbiont associations are subject to natural selection in the field. This selection may favour combinations where protection is conferred at a moderate cost. If this was indeed the case, the gain in aphid fitness from removing a natural infection with *H. defensa* should be lower on average than the loss of aphid fitness from an experimental infection. We could not confirm this hypothesis. Although there were genotype-by-genotype interactions between aphid clones and endosymbiont isolates as previously described (Chong and Moran 2016b; Vorburger and Gouskov 2011), we show that – over all *H. defensa* isolates – costs did not differ significantly between natural and experimental combinations. This seems to indicate that costs of *H. defensa* measured after introduction into a common genetic background are generally representative of costs in natural populations.

An important caveat is that we evaluated experimental infections in only two genetic backgrounds. If these two clones were not representative of the average susceptibility of *A. fabae* to infection with *H. defensa*, this might also have introduced a bias. Our conclusion further hinges on the assumption that curing aphids of *H. defensa* with an antibiotic (cefotaxime) does not have any lasting (multi-generational) effects on the aphids, e.g. by harming the primary endosymbiont *B. aphidicola*. Such an effect would lead to an underestimation of costs of natural infections as it would compromise the fitness of the cured

lines. Although we observed aphids that did not reproduce immediately after cefotaxime treatment, the comparison of naturally infected with cured and reinfected lines approximately 16 generations after antibiotic exposure indicates that antibiotics do not inflict persistent damage. Thus, experiments assessing costs of *H. defensa* in naturally infected and cured aphids should not suffer from bias if the aphids are allowed to recover from antibiotic exposure.

Surprising was the case of clone A06-402, for which the cured and re-infected line showed improved survival and reproduction compared to the original, naturally infected line. Potentially, this aphid clone might have possessed two strains of *H. defensa,* of which only one – which happened to be less costly – was transferred at reinfection. We do not consider this very likely as we have no evidence for a double infection, but with strain typing by Sanger sequencing it would be possible to miss a less common variant. Presence of several endosymbiont strains has been described for the endosymbiont *Regiella insecticola* in the pea aphid (Guyomar et al. 2018) but also in other organisms, such as ticks (Walter et al. 2016) or the bivalve *Solemya velum* (Russell and Cavanaugh 2017). Alternatively it is also possible that the naturally infected aphid line suffered from another, opportunistic infection at the time when the experiment took place (see below), especially given that its fitness was conspicuously low.

An additional concern is the stability of symbiont-induced phenotypes over time. How representative are the costs of infection with *H. defensa* estimated here of the costs at the time when the aphids and symbionts were collected in the field? After all, some isolates used in this experiment have been associated with their natural and experimental aphid partners for approximately a decade of laboratory culture. This is much longer than associations last in natural populations. Black bean aphids reproduce by cyclical parthenogenesis in central Europe (Sandrock et al. 2011), such that the sexual generation in autumn generates new combinations of host and symbiont genotypes every year. While the host's sexual recombination interrupts co-adaptation with the endosymbiont, it might also protect the host from exploitation by the symbiont (Stoy et al. 2020). In contrast to aphids in the wild, aphids in our long-term laboratory cultures do not reproduce sexually. Additionally, only a small number of adult aphids are used to found each subsequent generation. This likely allows drift to determine the aphid lines' evolution. Long-term association between host and endosymbiont and relaxed competition between hosts under the benign lab conditions might facilitate the evolution of selfish endosymbionts that become more and more costly (Bennett

and Moran 2015; Stoy et al. 2020). Yet, comparison of the costs that *H. defensa* inflicts on clones A06-405 and A06-407 over time revealed no clear trend towards increasing costs. Instead, some isolates seemed to follow independent trajectories towards increasing or decreasing costs in a way that is difficult to explain by environmental variation (different experimenters, different temperatures or potentially different humidity). Since decreased host fitness has already been connected to high endosymbiont titres in case of *B. aphidicola* (Chong and Moran 2016b), we compared fitness costs to *H. defensa* titre. Only the costs imposed by different isolates in 2019 correlated well with *H. defensa* titres assessed in 2018. Temporal variation in costs might thus be related to changes in symbiont population sizes, potentially caused by fluctuations in *H. defensa*'s virulence, its interaction with the host's immune system, or simply drift. Interestingly, however, no *H. defensa* isolate in our lab has ever lost its ability to protect aphids against parasitoids (see Cayetano and Vorburger 2013; Schmid et al. 2012; Vorburger and Rouchet 2016), despite complete absence of selection for *H. defensa*'s protective function in the long-term laboratory culture.

Another possible explanation for the temporal variation in costs is that our cultures may get infected occasionally by other pathogens, which could also affect lifespan or reduce reproduction. Since our current PCR primers target specific symbionts, we have no means of detecting invasion of pathogens such as viruses, fungi or even gut bacteria into the aphids' microbiome. The undetected presence of opportunistic pathogens would both explain the apparent changes in costs of host-isolate associations as well as the 'recovery' of the fitness of aphid clone A06-402 after antibiotic exposure and subsequent reinfection with its own *H. defensa* isolate.

In conclusion, our experiment indicates that in the absence of parasitoids, the fitness gain of losing a natural infection with *H. defensa* is comparable to the fitness loss from acquiring a new infection experimentally in black bean aphids. This indicates that assessing costs in a common host genetic background should be a valid strategy. However, the apparent instability of costs induced by different *H. defensa* isolates over time casts doubt on whether assessment in the lab after long-term laboratory culture is representative of the situation in the field. In the present case, both naturally infected and experimentally infected lines had been in long-term culture, hopefully precluding bias, but it has to be considered that the longer aphids are maintained in clonal cultures, the more the host-endosymbiont relationship may change. The reasons for the instability of *H. defensa*'s costs, as well as the apparent resistance to some or several antibiotics in *H. defensa* isolates of haplotype 1, warrants further investigation.

Declarations

Acknowledgments

We thank Silvan Rossbacher, Raffael Stegmayer and Nina Hafer-Hahmann for support with experimental procedures and Paula Rodriguez for the long-term maintenance of the insect lines. Jukka Jokela and Hugo Mathé-Hubert provided valuable input on the statistical analysis and study design.

Author's contributions

HK and CV designed the study and analysed the data. Aphids were cured by SJ and, in case of clone A06-101, by HK. Aphid reinfection and data collection was done by HK. The first draft of the manuscript was written by HK and all authors contributed to and commented subsequent versions. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Funding

This project has been funded by the Swiss National Science Foundation (Sinergia Grant No. CRSII3_154396).

Literature cited

Auguie B. 2015. gridExtra: Miscellaneous Functions for "Grid" Graphics.
Bennett GM, Moran NA. 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. Proceedings of the National Academy of Sciences. 112(33):10169.
Castañeda LE, Sandrock C, Vorburger C. 2010. Variation and covariation of life history traits in aphids are related to infection with the facultative bacterial endosymbiont *Hamiltonella defensa*. Biological Journal of the Linnean Society. 100(1):237-247.
Cayetano L, Rothacher L, Simon J-C, Vorburger C. 2015. Cheaper is not always worse: Strongly protective isolates of a defensive symbiont are less costly to the aphid host. Proceedings of the Royal Society B. 282(1799).
Cayetano L, Vorburger C. 2013. Genotype-by-genotype specificity remains robust to average temperature variation in an aphid/endosymbiont/parasitoid system. J Evol Biol. 26(7):1603-1610.
Chen D-Q, Montllor CB, Purcell AH. 2000. Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid, *A. kondoi*. Entomol Exp Appl. 95(3):315-323.
Chong RA, Moran NA. 2016. Intraspecific genetic variation in hosts affects regulation of obligate heritable symbionts. Proceedings of the National Academy of Sciences. 113(46):13114-13119.
Coeur d'Acier A, Sembène M, Audiot P, Rasplus JY. 2004. Polymorphic microsatellites loci in the black aphid, *Aphis fabae* Scopoli, 1763 (Hemiptera, *Aphididae*). Molecular Ecology Notes. 4(2):306-308.
Douglas AE. 1998. Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria *Buchnera*. Annual Review of Entomology. 43(1):17-37.
Dykstra HR, Weldon SR, Martinez AJ, White JA, Hopper KR, Heimpel GE, Asplen MK, Oliver KM. 2014. Factors limiting the spread of the protective symbiont *Hamiltonella defensa* in *Aphis craccivora* aphids. Appl Environ Microb. 80(18):5818-5827.
Ferrari J, Darby AC, Daniell TJ, Godfray HCJ, Douglas AE. 2004. Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. Ecological Entomology. 29(1):60-65.
Ferrari J, Scarborough CL, Godfray HCJ. 2007. Genetic variation in the effect of a facultative symbiont on host-plant use by pea aphids. Oecologia. 153(2):323-329.
Frossard J, Renaud O. 2019. Permutation tests for regression, ANOVA and comparison of signals: The permuco package.
Guyomar C, Legeai F, Jousselin E, Mougel C, Lemaitre C, Simon J-C. 2018. Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches. Microbiome. 6(1):181.
Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. P Natl Acad Sci USA. 108(7):2849-2854.
Leybourne DJ, Bos JIB, Valentine TA, Karley AJ. 2020. The price of protection: A defensive endosymbiont impairs nymph growth in the bird cherry-oat aphid, *Rhopalosiphum padi*. Insect science. 27(1):69-85.
Łukasik P, Dawid M, Ferrari J, Godfray HC. 2013a. The diversity and fitness effects of infection with facultative endosymbionts in the grain aphid, *Sitobion avenae*. Oecologia. 173(3):985-996.
Łukasik P, van Asch M, Guo H, Ferrari J, Charles J. Godfray H. 2013b. Unrelated facultative endosymbionts protect aphids against a fungal pathogen. Ecology Letters. 16(2):214-218.
McLean AHC, Hrček J, Parker BJ, Mathé-Hubert H, Kaech H, Paine C, Godfray HCJ. 2020. Multiple phenotypes conferred by a single insect symbiont are independent. Proceedings of the Royal Society B: Biological Sciences. 287(1929):20200562.
McLean AHC, van Asch M, Ferrari J, Godfray HCJ. 2011. Effects of bacterial secondary symbionts on host plant use in pea aphids. Proceedings Biological sciences. 278(1706):760-766.
Oliver KM, Campos J, Moran NA, Hunter MS. 2008. Population dynamics of defensive symbionts in aphids. Proceedings of the Royal Society B. 275(1632):293-299.

Oliver KM, Russell JA, Moran NA, Hunter MS. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proceedings of the National Academy of Sciences. 100(4):1803-1807.
Oliver KM, Smith AH, Russell JA. 2014. Defensive symbiosis in the real world – advancing ecological studies of heritable, protective bacteria in aphids and beyond. Functional Ecology. 28(2):341-355.
Parker BJ, Hrček J, McLean AHC, Godfray HCJ. 2017. Genotype specificity among hosts, pathogens, and beneficial microbes influences the strength of symbiont-mediated protection. Evolution; international journal of organic evolution. 71(5):1222-1231.
Poulin R, Keeney DB. 2008. Host specificity under molecular and experimental scrutiny. Trends Parasitol. 24(1):24-28.
R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rossbacher S, Vorburger C. 2020. Prior adaptation of parasitoids improves biological control of symbiont-protected pests. Evolutionary Applications. 13(8):1868-1876.
RStudio Team. 2020. RStudio: Integrated development for R. RStudio, PBC; Boston, MA.
Russell JA, Moran NA. 2006. Costs and benefits of symbiont infection in aphids: Variation among symbionts and across temperatures. Proceedings of the Royal Society B. 273(1586):603-610.
Russell SL, Cavanaugh CM. 2017. Intrahost genetic diversity of bacterial symbionts exhibits evidence of mixed infections and recombinant haplotypes. Molecular Biology and Evolution. 34(11):2747-2761.
Sandrock C, Razmjou J, Vorburger C. 2011. Climate effects on life cycle variation and population genetic architecture of the black bean aphid, *Aphis fabae*. Molecular Ecology. 20(19):4165-4181.
Sasaki T, Hayashi H, Ishikawa H. 1991. Growth and reproduction of the symbiotic and aposymbiotic pea aphids, *Acyrthosiphon pisum,* maintained on artificial diets. J Insect Physiol. 37(10):749-756.
Schmid M, Sieber R, Zimmermann Y-S, Vorburger C. 2012. Development, specificity and sublethal effects of symbiont-conferred resistance to parasitoids in aphids. Functional Ecology. 26(1):207-215.
Smith AH, Łukasik P, O'Connor MP, Lee A, Mayo G, Drott MT, Doll S, Tuttle R, Disciullo RA, Messina A et al. 2015. Patterns, causes and consequences of defensive microbiome dynamics across multiple scales. Molecular Ecology. 24(5):1135-1149.
Stoy KS, Gibson AK, Gerardo NM, Morran LT. 2020. A need to consider the evolutionary genetics of host–symbiont mutualisms. Journal of Evolutionary Biology. 00:1-13.
Sunnucks P, Hales DF. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: *Aphididae*). Mol Biol Evol. 13:510-524.
Therneau TM. 2020a. coxme: Mixed effects Cox models.
Therneau TM. 2020b. A Package for Survival Analysis in R.
Vorburger C. 2014. The evolutionary ecology of symbiont-conferred resistance to parasitoids in aphids. Insect science. 21(3):251-264.
Vorburger C, Ganesanandamoorthy P, Kwiatkowski M. 2013. Comparing constitutive and induced costs of symbiont-conferred resistance to parasitoids in aphids. Ecology and Evolution. 3(3):706-713.
Vorburger C, Gouskov A. 2011. Only helpful when required: A longevity cost of harbouring defensive symbionts. Journal of Evolutionary Biology. 24(7):1611-1617.
Vorburger C, Rouchet R. 2016. Are aphid parasitoids locally adapted to the prevalence of defensive symbionts in their hosts? Bmc Evol Biol. 16(1):271.
Vorburger C, Sandrock C, Gouskov A, Castañeda LE, Ferrari J. 2009. Genotypic variation and the role of defensive endosymbionts in an all-parthenogenetic host-parasitoid interaction. Evolution. 63(6):1439-1450.
Walter KS, Carpi G, Evans BR, Caccone A, Diuk-Wasser MA. 2016. Vectors as epidemiological sentinels: Patterns of within-tick *Borrelia burgdorferi* diversity. Plos Pathog. 12(7):e1005759.
Wickham H. 2007. Reshaping Data with the reshape Package. Journal of Statistical Software. 21(12):1-20.

Wickham H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.
Wickham H, Romain F, Lionel H, Müller K. 2020. dplyr: A Grammar of Data Manipulation.
Zytynska SE, Thighiouart K, Frago E. 2019. A meta-analysis on the benefits and costs of hosting secondary endosymbionts in sap-sucking insects. bioRxiv.563031.

Supplementary information

Supplementary Table 1 – Origin and experimental manipulation of aphid clones. Collection date, site and host plant for the aphid clones used for the experiment in their natural state, and date and type of treatment for experimentally manipulated aphids. *This table is provided in the electronic supplementary material.*

Supplementary Table 2 – Polymerase chain reaction conditions. Primers and cycling conditions for PCR reactions to confirm extraction success, presence of *H. defensa* and amplification of genes used for haplotype-typing.

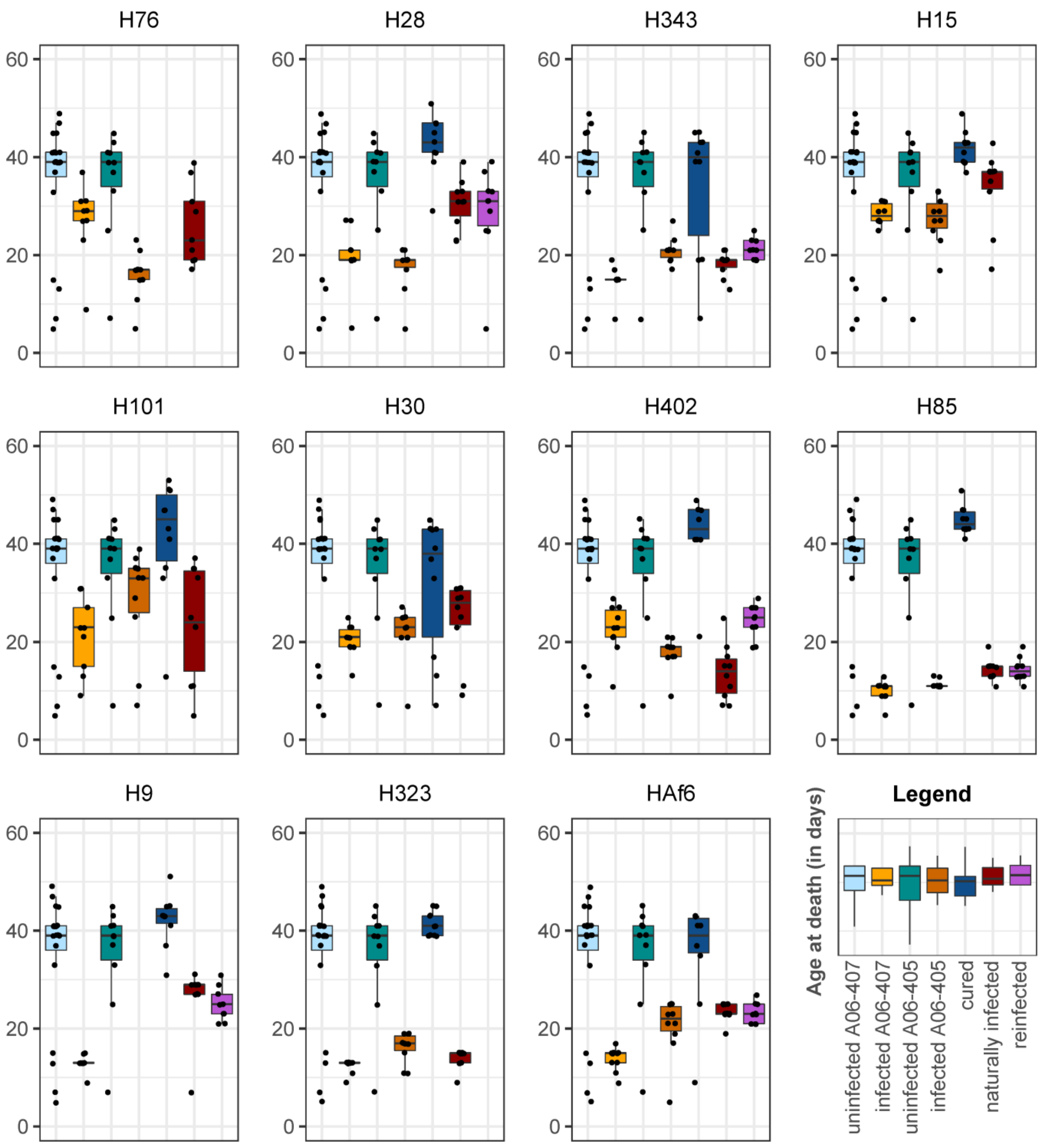
Primers targeting *Hamiltonella defensa*	
10F	5’-AGTTTGATCATGGCTCAGATTG-3’
T419/TO419	5’-AAATGGTATTSGCATTTATCG-3’

Primers targeting *Buchnera aphidicola*	
16SA1	5’-AGAGTTTGATCMTGGCTCAG-3’
Buch16SAfabR	5’-CTTCTTCGGGTAACGTCAAGAA-3’

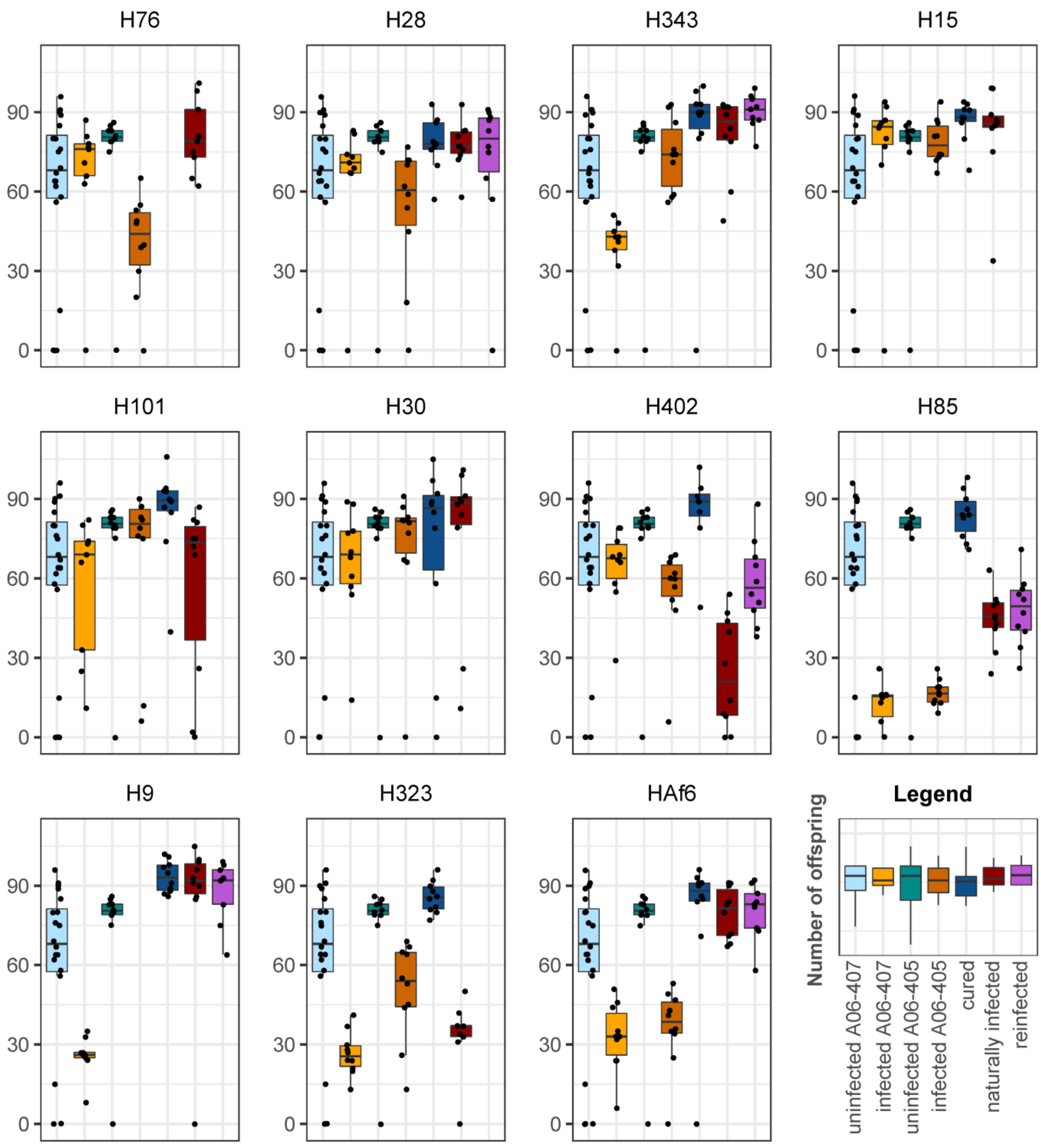
Primers for haplotype typing	
murE16F	5’-ACTAACGGGAAAACCACTAATAC-3’
murE936R	5’-TTGAGAATGTCAGCGGTAATC-3’
accD291F	5’-TTCTGGAGCACAAAAAGACAC-3’
accD832R	5’-AAGGTTCAGGTTGATGAGTCAG-3’

PCR program for endosymbionts		
Heat lid to 95°C		
95°C	3 min	
95°C	30 sec	10x
65-56°C	30 sec	
72°C	60 sec	
95°C	30 sec	25x
55°C	30 sec	
72°C	60 sec	
72°C	6 min	
10°C	∞	

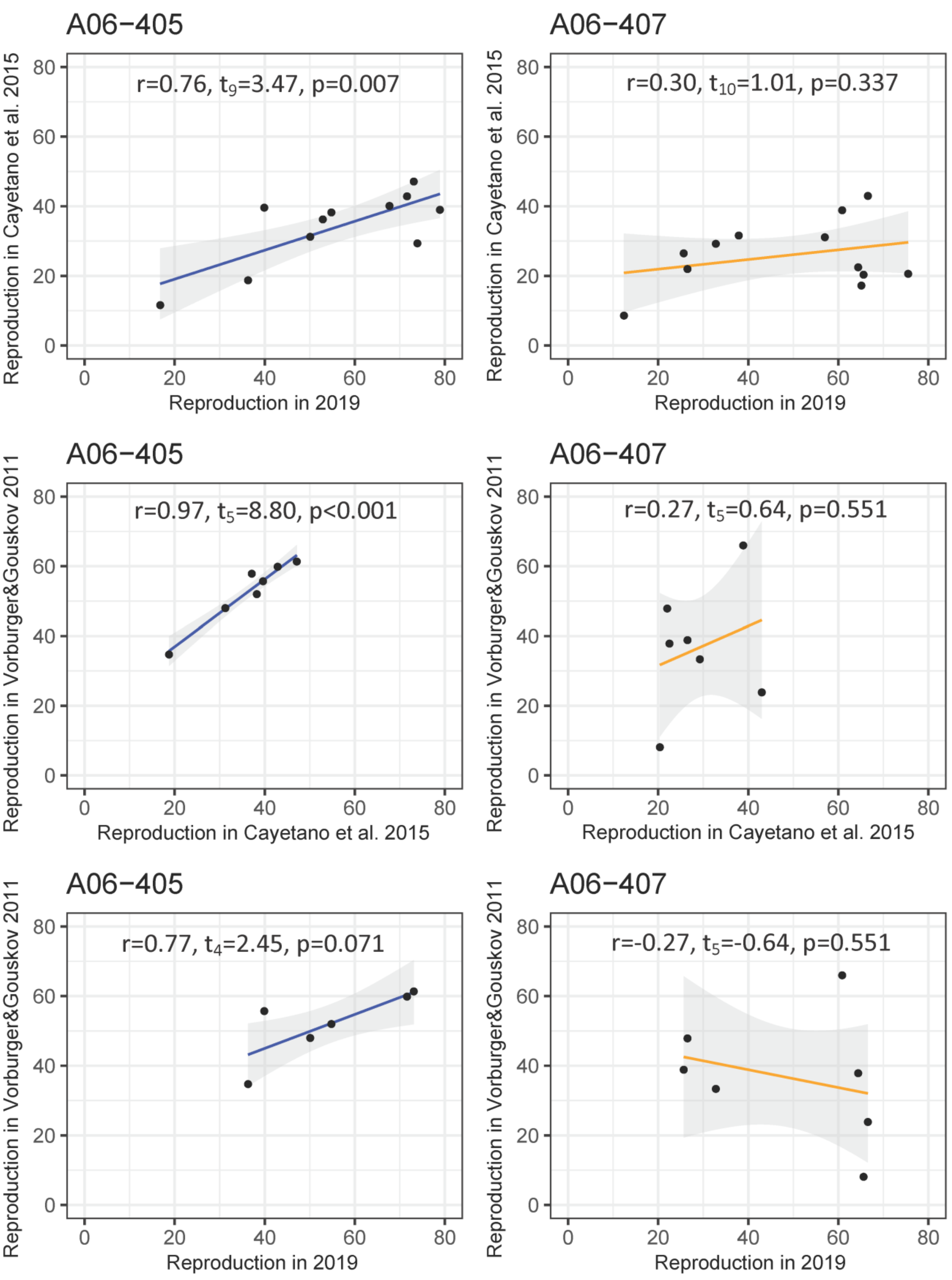
PCR program for haplotype testing		
Heat lid to 95°C		
94°C	2 min	
94°C	30 sec	11x
56-46°C	50 sec	
72°C	50 sec	
94°C	30 sec	25x
45°C	50 sec	
72°C	60 sec	
72°C	6 min	
4°C	∞	



Supplementary Figure 1 – Variation in lifespan depends on infection status. Age at death for all *A. fabae* clones and sublines infected with a specific *H. defensa* isolate. The legend is in the lower right corner. Light blue and turquoise indicate lifespan of the naturally uninfected aphid clones A06-407 ad A06-405, respectively. Dark yellow and orange indicate lifespan of A06-407 and A06-405, respectively, when infected with the *H. defensa* isolate. Dark blue, dark red and magenta indicate lifespan of the cured, naturally infected and reinfected clone that the *H. defensa* was associated with in nature.



Supplementary Figure 2 – Variation in lifetime reproduction depends on infection status. Number of offspring produced over their entire lifespan for all *A. fabae* clones and sublines infected with a specific *H. defensa* isolate. The legend is in the lower right corner. Light blue and turquoise indicate reproduction of the naturally uninfected aphid clones A06-407 ad A06-405, respectively. Dark yellow and orange indicate reproduction of A06-407 and A06-405, respectively, when infected with the *H. defensa* isolate. Dark blue, dark red and magenta indicate reproduction of the cured, naturally infected and reinfected clone that the *H. defensa* was associated with in nature.



Supplementary Figure 3 – Pearson's product-moment correlation of reproduction of aphid clones A06-405 and A06-407 with or without *H. defensa*-infection among three different experiments (Vorburger and Gouskov (2011), Cayetano et al. (2015) and this study).

General Discussion

The dispersal of a vertically transmitted, heritable endosymbiont hinges on its hosts' reproduction. Costly endosymbionts such as *H. defensa* or *Spiroplasma* can only spread in the host population, if they either manipulate their host's reproduction, or provide a fitness benefit (Doremus and Hunter 2020; Oliver et al. 2014). Repeatedly, endosymbionts of aphids have therefore evolved to protect their host against enemies (Guo et al. 2017). The evolution of such defensive symbioses is facilitated in vertically transmitted endosymbionts, since danger to the host also spells danger to the symbiont (Oliver et al. 2014; Vorburger and Perlman 2018).

In contrast to nutritional symbiosis, which is obligatory for the aphids, defensive symbionts of aphids have not gone to fixation in the host population. Theory predicts that intermediate frequency follows if the benefit of carrying a defensive symbiont turns into a liability in the absence of danger. Indeed, the frequency of *H. defensa* increases with parasitation pressure (Ives et al. 2020), and decreases in the absence of parasitoids (Dykstra et al. 2014; Oliver et al. 2008; Rossbacher and Vorburger 2020).

The cost of secondary endosymbiosis does not always manifest itself as reduced reproduction or lifespan in laboratory settings (Jamin and Vorburger 2019; Leonardo 2004; Łukasik et al. 2013b; Russell and Moran 2006), but the general pattern indicates that secondary endosymbionts are indeed costly – in aphids and in other insects (Cayetano et al. 2015; Chen et al. 2000; Herren et al. 2014; Oliver et al. 2008; Rossbacher and Vorburger 2020; Simon et al. 2011; Zytynska et al. 2019). Consistent with this pattern, we found costs of *Spiroplasma* on *A. pisum* (Mathé-Hubert et al. 2019), and *H. defensa* on *A. fabae* (chapter 3) on lifespan and reproduction. Yet, the results in chapter 3 also demonstrate the large variability of these costs – not only between different isolates of a symbiont, but also over time in an established symbiosis.

Validation of fitness assays

To measure the cost of secondary endosymbionts, we have to separate their fitness effects from the effects of the host genotype (Ferrari et al. 2007; Parker et al. 2017; Vorburger and Gouskov 2011). This can be achieved by either curing naturally infected hosts or by transfecting endosymbiont isolates into a common genetic background. While the first approach might lead to underestimation of costs if the antibiotic treatment reduces the host's

fitness, the second might lead to overestimation of costs if transfection leads to the creation of maladapted host-endosymbiont combinations.

In chapter 3 I validated both approaches. I showed that there is no lingering negative effect on host fitness 16 generations after antibiotic exposure, and that – on average – the fitness gain of removing *H. defensa* infections from naturally infected aphid clones is comparable to the fitness loss after transfecting *H. defensa* into naturally uninfected aphid clones. There is, however, the caveat that costs seemed to change during long-term clonal maintenance of host-endosymbiont associations.

This is not unexpected, as symbioses must not be seen as static but dynamic interactions, in which both partners are forced into a dance of coevolution (Bennett and Moran 2015; Chong and Moran 2016a; Stoy et al. 2020). The lack of a uniform direction of the changes, towards either reduced or increased cost, may be explained by the way these aphid lines have been maintained: in small populations with only a few aphids selected to found the next generation. Low effective population size and – potentially – reduced competition between aphids in the benign lab environment may lead to strong influence of drift on the host-endosymbiont evolutionary dynamics. Even if host-endosymbiont associations were under selection for high reproductive ability in the field, the signal of selection might have been lost from our natural host-endosymbiont association after nearly a decade of clonal maintenance. This would explain why in my experiment natural host-endosymbiont associations are no fitter than experimentally created associations.

Impact of a shortened life

But how bad are the costs measured in the lab really? Particularly a reduced lifespan does not necessarily translate to severe costs of *H. defensa* in the wild. Firstly, it is probably rare that an aphid reaches the proud age of 53 days observed in chapter 3 under natural conditions, given the danger of getting eaten, sick or parasitized. Secondly, offspring production wanes in older aphids: in my experiment, no aphid over the age of 35 days reproduced successfully. Thirdly, offspring produced earlier in life contribute more to an individual's fitness than late offspring in the explosively expanding aphid populations (Stearns 1992). There are compound measures of individual fitness that take not only reproduction and lifespan but also the timing of reproduction in account (Lenski and Service 1982). Their calculation is not trivial. As a rough approximation, I calculated the percentage of offspring that infected aphids produced compared to uninfected aphids. This rather simplistic approach assumes that infected lines

have the same rate of offspring production as cured lines. I first averaged the lifespan of all replicates of a naturally infected aphid clone. In each replicate of the corresponding cured aphid clone, I then calculated what percentage of its total offspring it had given birth to up to the age when infected aphids died. In two clones, the reduced lifespan severely curtailed offspring production, A06-85 with 62.1% and A06-323 with 66.2% of realised offspring (Supplementary Table 1). The reduction of lifespan clearly has a severe impact on offspring production in at least two natural host-endosymbiont associations.

Influence of transmission mode

It is not always known how costly endosymbionts offset their cost in order to survive in competition with 'cheaper' but more beneficial strains (Cayetano et al. 2015; Mathé-Hubert et al. 2019). On one hand, experiments such as described in chapter 3, which measure lifespan and lifetime reproduction, only assess single components of fitness. Other costs could arise from reduced nymphal growth (Leybourne et al. 2020) or behavioural change (Dion et al. 2011; Polin et al. 2014). On the other hand, there is a large range of traits that would have to be tested to fully assess an endosymbiont's benefits, such as deterrence of parasitoids (Łukasik et al. 2013a), changes in host plant use (Wagner et al. 2015) or resistance to heat stress (Montllor et al. 2002). Last but not least, costly endosymbionts may rely on an increased propensity for horizontal transmission to compensate their costs.

Indeed, isolates with higher titre tended to have more success at horizontal transmission (chapter 2). At the same time, however, titre alone was not enough to explain transmission success: The highly protective isolates of haplotype 1 (Cayetano et al. 2015; Dennis et al. 2017) transmit much more frequently than their titre suggests.

Mechanistic basic of *H. defensa'*s cost

In chapter 1, I explored the changes in aphid gene expression caused by infection with one of four different *H. defensa* strains. Even though I could not pinpoint which exact mechanisms cause the costs of these strains, the data indicated that different strains trigger distinctly different patterns of gene expression – and that *H. defensa* H76 indeed elicited little change in the aphid's gene expression. The experiment also provided ample motivation for further research: What does the set of strongly differentially regulated genes in aphids infected by *H. defensa* H402 do? Does *H. defensa* H85 indeed manipulate the aphid host's hemocytes,

and if yes, how? Does the increased APSE gene expression serve as a self-limiting factor in *H. defensa* H76 and H402, and what causes the different activity of these APSEs?

Two of these questions can and have to be addressed with qPCR assays. The interaction between *H. defensa* H85 and the aphid hemocytes must be confirmed by observing the expression of the two two hemocyte markers hemocytin and peroxidasin throughout the aphid's life. Additionally, changes in the number of hemocytes in the hemolymph of aphids with or without *H. defensa* can be recorded as described in Laughton et al. (2011). Based on the RNA-Seq data it should also be possible to design a qPCR assay measuring APSE expression, which in turn allows targeted selection on APSE activity (see Chrostek and Teixeira (2015) for a selection experiment in *Wolbachia*). If APSE activity can be experimentally manipulated, there is a good chance that the genes regulating its activity can be identified.

General conclusions

The complex interaction of the three players in this symbiosis – aphid, *H. defensa* and APSE – is far from unravelled. Chapter 1 indicated how difficult it is to understand their relationship and interpret the gene expression data. Yet the chapter also produced a set of hypotheses that now can be tested to further our understanding. In chapter 2, I showed that costly *H. defensa* isolates likely have hitherto unknown benefits that offset their cost, as their horizontal transmission success is not higher than that of seemingly more beneficial, highly protective isolates. Finally, in chapter 3 I validated the study of costs in a common genetic background or in cured aphids, while simultaneously raising questions about the evolutionary dynamics in long-term associations of hosts and endosymbionts.

Literature cited

Bennett GM, Moran NA. 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. Proceedings of the National Academy of Sciences. 112(33):10169.

Cayetano L, Rothacher L, Simon J-C, Vorburger C. 2015. Cheaper is not always worse: Strongly protective isolates of a defensive symbiont are less costly to the aphid host. Proceedings of the Royal Society B. 282(1799).

Chen D-Q, Montllor CB, Purcell AH. 2000. Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid, *A. kondoi*. Entomol Exp Appl. 95(3):315-323.

Chong RA, Moran NA. 2016. Intraspecific genetic variation in hosts affects regulation of obligate heritable symbionts. Proceedings of the National Academy of Sciences. 113(46):13114.

Chrostek E, Teixeira L. 2015. Mutualism breakdown by amplification of *Wolbachia* genes. PLOS Biology. 13(2):e1002065.

Dennis AB, Patel V, Oliver KM, Vorburger C. 2017. Parasitoid gene expression changes after adaptation to symbiont-protected hosts. Evolution. 71(11):2599-2617.

Dion E, Polin SE, Simon J-C, Outreman Y. 2011. Symbiont infection affects aphid defensive behaviours. Biology letters. 7(5):743-746.

Doremus MR, Hunter MS. 2020. Chapter Nine - The saboteur's tools: Common mechanistic themes across manipulative symbioses. In: Oliver KM, Russell JA, editors. Advances in Insect Physiology. Academic Press. p. 317-353.

Dykstra HR, Weldon SR, Martinez AJ, White JA, Hopper KR, Heimpel GE, Asplen MK, Oliver KM. 2014. Factors limiting the spread of the protective symbiont *Hamiltonella defensa* in *Aphis craccivora* aphids. Appl Environ Microb. 80(18):5818-5827.

Ferrari J, Scarborough CL, Godfray HCJ. 2007. Genetic variation in the effect of a facultative symbiont on host-plant use by pea aphids. Oecologia. 153(2):323-329.

Guo J, Hatt S, He K, Chen J, Francis F, Wang Z. 2017. Nine facultative endosymbionts in aphids. A review. Journal of Asia-Pacific Entomology. 20(3):794-801.

Herren JK, Paredes JC, Schüpfer F, Arafah K, Bulet P, Lemaitre B. 2014. Insect endosymbiont proliferation is limited by lipid availability. 3.

Ives AR, Barton BT, Penczykowski RM, Harmon JP, Kim KL, Oliver K, Radeloff VC. 2020. Self-perpetuating ecological–evolutionary dynamics in an agricultural host–parasite system. Nature Ecology & Evolution. 4(5):702-711.

Jamin AR, Vorburger C. 2019. Estimating costs of aphid resistance to parasitoids conferred by a protective strain of the bacterial endosymbiont *Regiella insecticola*. Entomol Exp Appl. 167(3):252-260.

Laughton AM, Garcia JR, Altincicek B, Strand MR, Gerardo NM. 2011. Characterisation of immune responses in the pea aphid, *Acyrthosiphon pisum*. J Insect Physiol. 57(6):830-839.

Lenski RE, Service PM. 1982. The statistical analysis of population growth rates calculated from schedules of survivorship and fecunidity. Ecology. 63(3):655-662.

Leonardo TE. 2004. Removal of a specialization-associated symbiont does not affect aphid fitness. Ecology Letters. 7(6):461-468.

Leybourne DJ, Bos JIB, Valentine TA, Karley AJ. 2020. The price of protection: A defensive endosymbiont impairs nymph growth in the bird cherry-oat aphid, *Rhopalosiphum padi*. Insect science. 27(1):69-85.

Łukasik P, Dawid M, Ferrari J, Godfray HC. 2013a. The diversity and fitness effects of infection with facultative endosymbionts in the grain aphid, *Sitobion avenae*. Oecologia. 173(3):985-996.

Łukasik P, van Asch M, Guo H, Ferrari J, Charles J. Godfray H. 2013b. Unrelated facultative endosymbionts protect aphids against a fungal pathogen. Ecology Letters. 16(2):214-218.

Mathé-Hubert H, Kaech H, Ganesanandamoorthy P, Vorburger C. 2019. Evolutionary costs and benefits of infection with diverse strains of *Spiroplasma* in pea aphids. Evolution. 73(7):1466-1481.

Montllor CB, Maxmen A, Purcell AH. 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. Ecological Entomology. 27(2):189-195.

Oliver KM, Campos J, Moran NA, Hunter MS. 2008. Population dynamics of defensive symbionts in aphids. Proceedings of the Royal Society B. 275(1632):293-299.
Oliver KM, Smith AH, Russell JA. 2014. Defensive symbiosis in the real world – advancing ecological studies of heritable, protective bacteria in aphids and beyond. Functional Ecology. 28(2):341-355.
Parker BJ, Hrček J, McLean AHC, Godfray HCJ. 2017. Genotype specificity among hosts, pathogens, and beneficial microbes influences the strength of symbiont-mediated protection. Evolution; international journal of organic evolution. 71(5):1222-1231.
Polin S, Simon J-C, Outreman Y. 2014. An ecological cost associated with protective symbionts of aphids. Ecology and Evolution. 4(6):836-840.
Rossbacher S, Vorburger C. 2020. Prior adaptation of parasitoids improves biological control of symbiont-protected pests. Evolutionary Applications. 13(8):1868-1876.
Russell JA, Moran NA. 2006. Costs and benefits of symbiont infection in aphids: Variation among symbionts and across temperatures. Proceedings of the Royal Society B. 273(1586):603-610.
Simon J-C, Boutin S, Tsuchida T, Koga R, Le Gallic J-F, Frantz A, Outreman Y, Fukatsu T. 2011. Facultative Symbiont Infections Affect Aphid Reproduction. PLoS ONE. 6(7):e21831.
Stearns SC. 1992. The evolution of life histories. New York: Oxford University Press.
Stoy KS, Gibson AK, Gerardo NM, Morran LT. 2020. A need to consider the evolutionary genetics of host–symbiont mutualisms. Journal of Evolutionary Biology. 00:1-13.
Vorburger C, Gouskov A. 2011. Only helpful when required: A longevity cost of harbouring defensive symbionts. Journal of Evolutionary Biology. 24(7):1611-1617.
Vorburger C, Perlman SJ. 2018. The role of defensive symbionts in host–parasite coevolution. Biological Reviews. 93(4):1747-1764.
Wagner SM, Martinez AJ, Ruan Y-M, Kim KL, Lenhart PA, Dehnel AC, Oliver KM, White JA. 2015. Facultative endosymbionts mediate dietary breadth in a polyphagous herbivore. Functional Ecology. 29(11):1402-1410.
Zytynska SE, Thighiouart K, Frago E. 2019. A meta-analysis on the benefits and costs of hosting secondary endosymbionts in sap-sucking insects. bioRxiv.563031.

Supplementary information

Supplementary Table 1 – Impact of shortened lifetime on the percentage of realised offspring produced by *H. defensa*-infected aphids compared to uninfected aphids of the same genotype.

Aphid clone	Average lifespan when infected	Average percentage of offspring realised compared to cured	Standard error
A06-101	23.8	94.4	1.8
A06-28	30.6	100.0	0.0
A06-28 reinfected	28.8	99.6	0.3
A06-30	24.6	99.7	0.3
A06-323	13.9	66.2	2.9
A06-343	18.2	97.6	0.8
A06-343 reinfected	21.2	100.0	0.0
A06-402	13.8	50.4	5.9
A06-402 reinfected	24.4	93.3	2.6
A06-9	26.4	100.0	0.0
A06-9 reinfected	25.0	97.5	1.4
A06-Af6	23.4	99.2	0.4
A06-Af6 reinfected	23.2	99.2	0.4
A06-15	33.8	100.0	0.0
A06-85	14.4	62.1	4.7
A06-85 reinfected	14.4	62.1	4.7

Acknowledgements

When doing research we build on the past. On the hard work and dedication of other scientists, all the way back through history. But the past is not enough: We also need a nurturing environment in the present.

I was lucky to share my journey with fantastic peers. A group of empathic, witty and kind Ph. D. candidates who met most Monday mornings to eat croissants, share woes and celebrate whoever published a paper with Prosecco (yes, at 8:15 am). Their support kept me sane, their honesty saved me from imposter syndrome, and the time spent with them was simply great fun.

I had the privilege of having an advisor who supported me unconditionally. Not only did he take a risk by allowing me to embrace science journalism while finishing my doctorate, he also went above and beyond to help me push through the last stretch of my thesis in time. I am eternally grateful for his dedication, trust and kindness.

This kindness of his is reflected in the people that are attracted to his lab – wonderful technicians who kept the entire lab running with their dedication and willingness to help however short term a request, as well as postdocs who always had time and patience for me; for inspiring feedback, solid advice and a pep talk when I needed it most.

Last but not least I have been surrounded by amazing people. Dear friends who brightened up even the most stressful days with a message, a call or a shared lunch break. An aunt whose council has been indispensable, a husband who went through this all with me and for me, parents who I know I can always rely on, and grandparents who always hoped to get to see me defend my thesis.

To all of you: Thank you.

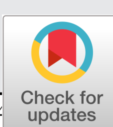
Appendix chapter 1

Hugo Mathé-Hubert, Heidi Kaech, Pravin Ganesanandamoorthy, and Christoph Vorburger

Evolutionary costs and benefits of infection with diverse strains of *Spiroplasma* in pea aphids

(2019; Evolution 73: 1466-1481; https://doi.org/10.1111/evo.13740)

ORIGINAL ARTICLE



doi:10.1111/evo.1374

Evolutionary costs and benefits of infection with diverse strains of *Spiroplasma* in pea aphids*

Hugo Mathé-Hubert,[1,2,3] **Heidi Kaech,**[1,4] **Pravin Ganesanandamoorthy,**[1] **and Christoph Vorburger**[1,4]

[1]*Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133 8600 Dübendorf, Switzerland*

[2]*Current Address: LIEC UMR 7360, Université de Lorraine and CNRS, Metz, France*

[3]*E-mail: hugomh@gmx.fr*

[4]*Institute of Integrative Biology, Department of Environmental Systems Science, ETH Zürich, Universitätsstrasse 16 8092 Zürich, Switzerland*

Received October 31, 2018
Accepted March 28, 2019

The heritable endosymbiont *Spiroplasma* infects many insects and has repeatedly evolved the ability to protect its hosts against different parasites. Defenses do not come for free to the host, and theory predicts that more costly symbionts need to provide stronger benefits to persist in host populations. We investigated the costs and benefits of *Spiroplasma* infections in pea aphids (*Acyrthosiphon pisum*), testing 12 bacterial strains from three different clades. Virtually all strains decreased aphid lifespan and reproduction, but only two had a (weak) protective effect against the parasitoid *Aphidius ervi*, an important natural enemy of pea aphids. *Spiroplasma*-induced fitness costs were variable, with strains from the most slowly evolving clade reaching higher titers and curtailing aphid lifespan more strongly than other strains. Some *Spiroplasma* strains shared their host with a second endosymbiont, *Regiella insecticola*. Although the result of an unfortunate handling error, these co-infections proved instructive, because they showed that the cost of infection with *Spiroplasma* may be attenuated in the presence of *Regiella*. These results suggest that mechanisms other than protection against *A. ervi* maintain pea aphid infections with diverse strains of *Spiroplasma*, and that studying them in isolation will not provide a complete picture of their effects on host fitness.

KEY WORDS: ***Acyrthosiphon pisum*, cost of infection, defensive symbiosis, facultative secondary symbionts, lifespan, parasitoid.**

Introduction

Microbial endosymbionts of eukaryotes are ubiquitous, and have often become heritable through the evolution of mother-to-offspring transmission. Large-scale screens for symbionts like *Wolbachia* or *Cardinium* suggest that the majority of arthropod species are likely to carry heritable infections with endosymbionts (Zchori-Fein and Perlman 2004; Hilgenboecker et al. 2008). Microbial symbionts may provide their hosts with essential nutrients, especially in species with very imbalanced diets such as blood feeders like the tsetse fly (Chen et al. 1999) or phloem feeders like aphids (Douglas 1998). Some of these symbioses are ancient and have evolved to the point that the host is unable to survive without its bacterial partner, which is referred to as an obligate symbiont (Wernegreen 2002; Moran et al. 2008). Other endosymbionts are facultative associates for the host and not strictly required for host survival. These are referred to as secondary symbionts. A single arthropod species can host multiple species of secondary symbionts, but each symbiont typically infects only a part of the host population (e.g., Chiel et al. 2007; Ferrari et al. 2012). Additional variation may be present within symbionts. A secondary symbiont species infecting a particular host species often comprises multiple distinguishable strains (Raychoudhury et al. 2009; Ferrari et al. 2012; Russell et al. 2013). Explaining the evolutionary persistence and the high diversity of secondary symbionts in host populations requires an understanding of how different symbionts counterbalance the costs they impose on their host (Heath and Stinchcombe 2014).

One way for maternally transmitted symbionts to spread in a host population is to manipulate the host's reproduction in a way

*This article corresponds to El-Deeb, O. 2019. Digest: Fitness costs of *Spiroplasma* infection in pea aphids. Evolution. https://doi.org/10.1111/evo.13763.

© 2019 The Author(s). *Evolution* © 2019 The Society for the Study of Evolution.

that favors symbiont transmission. Reproductive manipulation has evolved repeatedly in endosymbiotic bacteria like *Wolbachia*, *Arsenophonus*, *Cardinium*, *Rickettsia*, or *Spiroplasma* (Duron et al. 2008). It can act via the induction of cytoplasmic incompatibility, male-killing, parthenogenesis, or the feminization of genetically male offspring (Werren et al. 2008).

In addition to reproductive manipulation, heritable symbionts can spread if they provide their host with an evolutionary benefit. This strategy is not mutually exclusive with reproductive manipulation. An important class of evolutionary benefits that has evolved repeatedly is protection against natural enemies, that is, defensive symbiosis (Oliver and Moran 2009; McLean 2019). Multiple species of secondary symbionts increase the resistance of aphids against parasitoid wasps and pathogenic fungi (Oliver et al. 2003; Scarborough et al. 2005; Vorburger et al. 2010; Łukasik et al. 2013), certain strains of *Spiroplasma* can protect flies against parasitoid wasps or parasitic nematodes (Jaenike et al. 2010; Xie et al. 2010; Paredes et al. 2016), and *Wolbachia* can reduce viral infection in flies and other insects (Hedges et al. 2008; Teixeira et al. 2008; Bian et al. 2010). So why do these seemingly beneficial symbionts not go to fixation in host populations?

Most general explanations assume trade-offs between the benefits provided by the symbiont and the costs associated with its possession, acting in combination with environmental heterogeneity. For example, the secondary symbiont *Hamiltonella defensa* (Moran and Russell 2005) can protect different aphid species against parasitism (Oliver et al. 2003; Schmid et al. 2012; Asplen et al. 2014), but *H. defensa* is selected against in the absence of parasitoids (Oliver et al. 2008), possibly because of the reductions in host lifespan and lifetime reproduction or in nymphal growth it induces (Vorburger and Gouskov 2011; Leybourne et al. 2018). Temporal and spatial variation in the risk of parasitism may thus maintain coexistence between infected and uninfected hosts. Similarly, species and strain diversity may partly be explained by unequal effects against different natural enemies. For *H. defensa*, several studies have shown that protection of aphids against parasitoid wasps can be highly specific (reviewed in Vorburger 2014). A given strain of *H. defensa* can provide effective protection against some parasitoid species but not against others (Asplen et al. 2014; Cayetano and Vorburger 2014; McLean and Godfray, 2015, 2017; Martinez et al. 2016), and this specificity can even extend to interactions within species. In black bean aphids (*Aphis fabae*), particular isolates of *H. defensa* protect strongly against some parasitoid genotypes but not or only weakly against other parasitoid genotypes, leading to strong genotype-by-genotype interactions between parasitoids and the hosts' defensive symbionts (Schmid et al. 2012; Cayetano and Vorburger 2013; Vorburger and Rouchet 2016). Similar genotype-specificity is observed in the interaction between the fungal pathogen *Pandora neoaphidis* and the secondary symbiont *Regiella insecticola*, which protects pea aphids (*Acyrthosiphon pisum*) against fungal infection (Parker et al. 2017). Variation in the local parasitoid and pathogen community may thus select for different secondary symbionts, and genotype-by-genotype specificity may further maintain strain variation via negative frequency-dependent selection (Kwiatkowski et al. 2012; Heath and Stinchcombe 2014).

A promising system to investigate the evolutionary maintenance of symbiont strain diversity are bacteria of the genus *Spiroplasma*. These helical, cell wall-less bacteria belong to the class Mollicutes within the phylum Firmicutes (Gasparich et al. 2004). *Spiroplasma* bacteria are generally associated with arthropods, but they differ widely in their modes of transmission and their phenotypic effects on the hosts. Some are virulent, horizontally transmitted pathogens of insects and crustaceans that cause problems in apiculture and aquaculture (Clark et al. 1985; Wang et al. 2005), some are damaging plant pathogens that are vectored by phloem-feeding insects (Bové et al. 2003), and many are vertically transmitted endosymbionts (Williamson et al. 1998; Watts et al. 2009). It is estimated that between 5% and 10% of insects carry heritable infections with *Spiroplasma* (Duron et al. 2008). Similar to other heritable endosymbionts, some *Spiroplasma* have evolved the ability to defend their hosts against other infections (Ballinger and Perlman 2018). For example, the male killing strain MSRO of *S. poulsonii* protects *Drosophila melanogaster* against parasitoid wasps (Xie et al. 2014; Paredes et al. 2016), illustrating that reproductive manipulation and protection are not mutually exclusive strategies of symbionts to spread in host populations. In the fungus-feeding *D. neotestacea*, infection with *Spiroplasma* induces tolerance to the parasitic nematode *Howardula aoronymphium* (Jaenike et al. 2010). In pea aphids, *Spiroplasma* has been shown to protect against fungal infections (Łukasik et al. 2013), and there is evidence for male-killing by at least one strain (Simon et al. 2011).

Spiroplasma infecting European pea aphids are subdivided into at least three clades that are similarly abundant in aphids feeding on different host plants, but share their hosts with different symbiont communities and have a different rate of molecular evolution, suggesting their maintenance in pea aphids might rely on different eco-evolutionary strategies (Mathé-Hubert et al. 2018). Here, we provide insights in the ecology and evolution of these three clades. We investigate if protection against the pea aphid's main parasitoid *Aphidius ervi* might contribute to the evolutionary persistence of *Spiroplasma* in this species. Twelve *Spiroplasma* strains, evenly spread across the three clades, were tested for their ability to protect against three different lines of the parasitoid *A. ervi*, and we estimated *Spiroplasma* density in 10- and 20-day-old aphids, as well as *Spiroplasma*'s effects on aphid fitness in the absence of parasitoids. Although two of the 12 *Spiroplasma* strains reduced aphid parasitism by at least one of three parasitoid

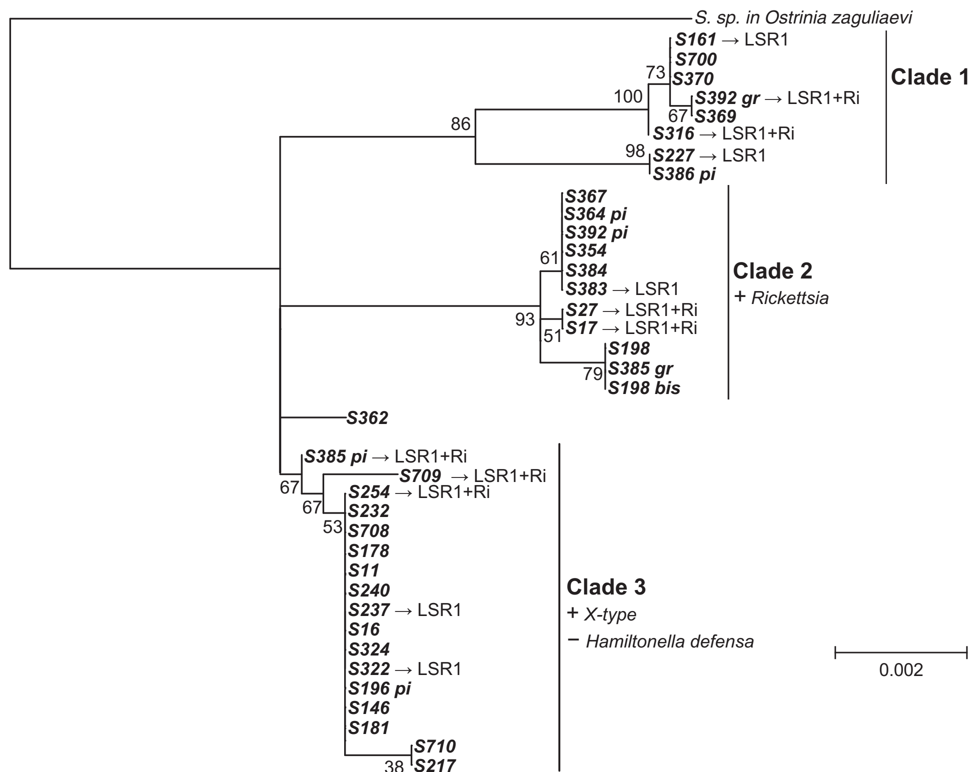


Figure 1. ***Spiroplasma* phylogeny. Phylogeny modified from Mathé-Hubert et al. (2018). Strains selected for the phenotyping experiments are followed by an arrow indicating whether they were transfected into pea aphid clone LSR1 containing *R. insecticola* (→ LSR1+Ri) or not (→ LSR1). The clade to which the *Spiroplasma* strain belongs and the other symbionts with which strains of that clade are typically associated (+) or not associated (–) is indicated on the right. Values close to the nodes are bootstrap values. The scale bar indicates the substitution rates.**

lines, there was no global effect of *Spiroplasma* on the parasitism success. All *Spiroplasma* strains curtailed aphid lifespan and lifetime reproduction to various extents and the benefit provided by the two protective *Spiroplasma* strains is unlikely to counterbalance their cost. This suggests that *Spiroplasma* infection in pea aphids is maintained by another mechanism than the protection against *A. ervi*. An analysis of phylogenetic signal in the phenotypic data further revealed that the most slowly evolving of the three *Spiroplasma* clades attains the highest titer in aphids and reduces aphid lifespan more strongly than the other clades.

Material and Methods

INSECT LINES

To investigate phenotypic effects of *Spiroplasma* infections, we used the European field survey and the phylogeny of *Spiroplasma* from pea aphids reported in Mathé-Hubert et al. (2018) to select 12 strains that are well spread across the phylogeny (Fig. 1). To control for the effect of aphid genotype, the selected *Spiroplasma* strains were transfected from their original host clones (the donors) into a common recipient clone called LSR1. This clone was originally collected in a field of alfalfa (*Medicago sativa*) near Ithaca, New York, in 1998 (Caillaud et al. 2002), and its genome has been sequenced for the pea aphid genome project (The International Aphid Genomics Consortium 2010). Four *Spiroplasma* strains were transfected into LSR1 at the University of Oxford, U.K., and kindly provided to us by Ailsa McLean. The remaining transfections were carried out in our laboratory at Eawag, Switzerland. Before transfection of *Spiroplasma* with a microinjection pump (FemtoJet, Eppendorf) as described by Vorburger et al. (2010), the donor aphids were cured from all other secondary symbionts by feeding them on a mixture of antibiotics as described in McLean et al. (2011). For strain S383, this protocol failed to remove a co-infection with *Hamiltonella defensa* in the donor clone. We thus merged the curing and transfection step by injecting recipients with a small amount of a 20 mg/mL solution of the antibiotic cefotaxime, using a needle that was immersed into the donor's hemolymph prior to injection. This procedure succeeded in transmitting just *Spiroplasma* to the recipient clone. Depending on the strains, the transfections happened between 10 and ~150 generations before the experiments.

Although prior to transfections we had reconfirmed the genotypes and the secondary symbiont infections of the donors and the

recipient clone with microsatellites and diagnostic PCRs, respectively, a handling error must have occurred between these checks and the actual transfections, such that we used a *R. insecticola*-infected sub-line of clone LSR1 (LSR+Ri) as recipient rather than the sub-line without any secondary symbionts. As a consequence, seven of the 12 newly transfected sub-lines carried a co-infection with *R. insecticola* in addition to the different *Spiroplasma* strains. Only sub-line LSR1+S383 (presumably due to the simultaneous injection of an antibiotic-see above), the four sub-lines provided by the University of Oxford, and the secondary endosymbiont-free control did not carry *R. insecticola*. Figure 1 summarizes the infection status of each sub-line. That the *R. insecticola*-infected sublines indeed belonged to clone LSR1 was confirmed by microsatellite genotyping, and sequencing of five bacterial genes (*accD*, *gyrB*, *murE*, *recJ*, and *rpoS*; Henry et al. 2013) identified the co-infecting *R. insecticola* as a strain previously shown to provide no protection against *A. ervi* in pea aphids (Oliver et al. 2003; Hansen et al. 2012). Because we discovered this error only after all phenotyping experiments had been completed, we had to account statistically for the presence of *R. insecticola* during data analysis (see below).

ACCOUNTING STATISTICALLY FOR CO-INFECTIONS WITH *R. insecticola*

For the three experiments described hereafter, we handled the presence of *R. insecticola* according to the following logic: We estimated the average effect of *R. insecticola* on each trait we analyzed and then used this estimate as an offset to correct for its presence in the coinfected sublines. Specifically, we first fit a "*Regiella*" model devised to estimate the average effect of *R. insecticola* in the presence of a *Spiroplasma* strain. In addition to the variables specific to each experiment (described in the corresponding sections), this model contains two dummy variables as fixed effects describing the presence (1) or absence (0) of *Spiroplasma* and *R. insecticola* (variables S_i and R_i, respectively) and a random interaction between the aphid subline (*SUB*) and the fixed effect *S*. This random effect follows a normal distribution of mean zero and standard deviation σ. Mathematically, this gives

$$Y_i = Int + \alpha \times R_i + \beta \times S_i + S_i \times SUB + e_i;\; SUB \sim N(0, \sigma) \quad (1)$$

where Y_i is the transformed explained variable, e_i are the residuals estimated by the models together with the coefficients of the fixed effects (α and β) and the standard deviation (σ). Because we used dummy variables, the intercept of the model (*Int*) is the mean of the control sub-line containing neither *Spiroplasma* nor *R. insecticola*. The coefficients α and β are the estimated mean effects of *R. insecticola* and *Spiroplasma*, and the random interaction between the sub-line and *S* accounts for the heterogeneity induced by the different *Spiroplasma* strains. This estimation of the effect of *R. insecticola* assumes that on average the *Spiroplasma* strains that are alone have the same effect as the *Spiroplama* strains that are with *R. insecticola*. The estimated effects of *R. insecticola* (coefficient α in eq. 1) is then used to construct an offset (Hutchinson and Holtman 2005) for the second "*Spiroplasma*" model estimating the effect of each *Spiroplasma* strain. This offset takes the value α when *R. insecticola* is present and 0 when it is absent. The "*Spiroplasma*" model contains the aphid sub-line as a fixed effect. Thus, for the sub-lines not containing *R. insecticola*, there is no offset and each coefficient describes the effect of the sub-line's *Spiroplasma* strain, and for the sub-lines containing *R. insecticola*, the estimated effect of *R. insecticola* in the presence of *Spiroplasma* is absorbed by the offset, and each coefficient describes the effect of the corresponding *Spiroplasma* strain plus its eventual interaction with *R. insecticola*.

EXPERIMENT 1: EFFECT OF *SPIROPLASMA* ON *A. ervi* PARASITISM

We investigated the effect of the 12 *Spiroplasma* strains on the parasitism success of three different lines of the parasitoid wasp *A. ervi* (lines "B," "D," and "K"). We established the line "D" using wasps sampled in July 2015 at two sites in southern Germany during the field survey reported in Mathé-Hubert et al. (2018). This wasp line has been maintained in the laboratory for approximately 40 generations prior to the experiment. The two other *A. ervi* lines "K" and "B" were commercially supplied by the biocontrol companies Koppert (Berkel en Rodenrijs, the Netherlands) and Biobest (Westerlo, Belgium), and were reared in the laboratory for one and two generations before the experiment, respectively. We used three different lines of parasitoids to increase our chances of detecting any protective effects of *Spiroplasma*, since previous studies on another bacterial endosymbiont, *H. defensa*, have shown that the protection afforded by the symbiont can depend on the parasitoid's genotype (e.g., Schmid et al. 2012; Cayetano and Vorburger 2013). All wasps were bred on the same pea aphid clone (lab ID A06-01) that was free of protective endosymbionts and different from the clone used in experiments (LSR1).

Parasitism success was measured using a factorial design in which the 13 aphid sub-lines (12 *Spiroplasma*-infected sub-lines plus uninfected control) were exposed to all three parasitoid lines in six randomized complete blocks. To prevent maternal effects carried over from the aphid stock cultures influencing our results, each of the 234 replicates (13 aphid sub-lines × 3 wasp lines × 6 replicates) was reared independently on seedlings of broad bean (*Vicia faba*) for one generation before individuals of the second generation were tested. To start the test generation, five adults from each replicate were used to obtain age-synchronized offspring born within 24 h. At the age of 2–3 days, 20 nymphs

per replicate were placed on a new plant and exposed to a single female wasp (~2 days old) for 5 h. Because a few aphid nymphs were harmed during the exposure to wasps, the number of nymphs alive one day after the exposure was recorded. The proportion of these surviving nymphs that were successfully parasitized and transformed into mummies (parasitoid pupae within the dead aphid's exoskeleton) was recorded 11 days after exposure to parasitoids. The proportion of mummies from which adult wasps had emerged successfully (proportion emerged) was recorded 20 days after exposure. We conducted this experiment at 22°C under a 16-h photoperiod.

For each of the two variables, proportion mummified and proportion emerged, we fitted the "*Regiella*" and "*Spiroplasma*" models as described above. Both models additionally contained the wasp line as a fixed effect as well as its interaction with the dummy variables "*R*" and "*S*" for the model "*Regiella*" and with the aphid sub-line for the model "*Spiroplasma*." Both models also contained the random variable "Block."

If for the "*Spiroplasma*" model the wasp line × aphid sub-line interaction was significant, we re-fitted one model per wasp line to test for overall variation among aphid sub-lines and to assess the effect of each *Spiroplasma* strain using a Student's *t*-test. These tests compare each *Spiroplasma*-infected sub-line to the uninfected control sub-line by assessing the significance of the coefficients of the variable "aphid sub-line." We then used the package "multcomp" (Hothorn et al. 2008) to assess for each of the models fitted to one wasp line that *Spiroplasma* strains had a significant effect after accounting for multiple testing. When the wasp line × aphid sub-line interaction was not significant, we refitted the model without the interaction to test for the effect of each *Spiroplasma* strain.

Since the explained variables were proportions, we first fitted them using binomial GLMMs ("lme4" R package; Bates et al. 2014), which were strongly overdispersed. The attempt to mitigate overdispersion with the "observation level random effect" approach (Harrison 2015) resulted in severe underdispersion. Thus, we fitted LMMs to the logit transformed proportions (Warton and Hui 2011). To assess the significance of the main effects, we used the "mixed" function of the "afex" R package (version 0.18) to perform an *F*-test with the Kenward–Roger approximation for degrees of freedom (Halekoh and Højsgaard 2014). All statistical analyses were performed using the software R (version 3.5.2).

EXPERIMENT 2: FITNESS COST OF *SPIROPLASMA*

We assessed the fitness cost of *Spiroplasma* strains by measuring their effects on several life-history traits of their host using the surplus of nymphs produced in experiment 1: For three of the six blocks, each containing three replicates of every aphid sub-line, we kept all leftover nymphs until they were 6 days old. Then, for each of the 117 replicates (13 aphid sub-lines × 9 replicates organized into 3 blocks), we selected two young aphids for the life table experiment. In 40% of the cases, one of the two aphids developed wings. They were excluded from the experiment. The 199 wingless aphids were raised individually on broad bean seedlings until their death. Every week, we moved the aphids to a new 9-day-old plant, and recorded the number of offspring they had produced on the former plant. We recorded the survival of the monitored aphids three times a week. The experiment was carried out at 18°C and under a 16-h photoperiod.

We used the life table data to estimate four fitness-related life history traits. The first two are lifetime reproduction (total number of offspring) and lifespan. We also computed the mean reproductive age of each aphid (age of mother at each birth, averaged across all offspring births). In comparison to the lifetime reproduction, the mean reproductive age accounts for the fact that two genotypes with the same lifetime reproduction could have different fitness if one of them produced its offspring earlier than the other. The fourth variable was the intrinsic growth rate, that is, the constant *r* in the equation describing population growth in an unlimited environment: $N_t = N_0\, e^{rt}$. The procedure to calculate it is described in Birch (1948). This variable combines the information of the number of offspring and of the age of the mother when the offspring are produced.

To each of these four fitness-related variables, we fitted the '*Regiella*' and '*Spiroplasma*' models. Both models also included the random variables block and replicate, the latter accounting for the non-independence of the two individuals taken from the same colony of experiment 1. The test procedure for these four variables is the same as described for experiment 1, except that a box-cox transformation was used to achieve normality of residuals and homoscedasticity instead of the logit function. For the survival data, we used the "coxme" R package (version 2.2-5) to fit a cox model (Therneau 2015a). For this survival analysis, we checked the assumption of proportional hazard using the "cox.zph" function of the package "survival" (Therneau 2015b; version 2.43-3) and the "survplot" function of the package "rms" (Harrell 2017; version 5.1-2), with the argument "loglog" set to true. As in experiment 1, this model assessed the overall variation among aphid sub-lines and compared each *Spiroplasma*-infected sub-line to the uninfected control.

EXPERIMENT 3: VARIATION IN *SPIROPLASMA* DENSITY

The density of *Spiroplasma* within its host may influence both the cost *Spiroplasma* inflicts on the aphid and the parasitism by *A. ervi*. Thus, we measured the density of *Spiroplasma* in 10- and 20-day-old aphids using quantitative PCR (qPCR). For each combination of age and strain, we measured five biological replicates, each consisting of a pool of three aphids that were

reared on a 9-day-old plant, a different plant being used for each biological replicate. The biological replicates were reared within a single tray on randomized positions. DNA was extracted using either the Qiagen "DNeasy 96 Blood & Tissue Kit" (extraction in plates; $N = 104$ samples) or the Qiagen "DNeasy Blood & Tissue Kit" (extraction in tubes; $N = 16$ samples) after the aphids had been crushed by shaking them 30 times per second for 40 sec with two glass marbles of 2 mm Ø on a bead mill (TissueLyser II, Qiagen). These extractions typically yield approximately 5 μg of DNA in 200 μL.

For each pool of three aphids, the number of *Spiroplasma* and aphid gene copies were estimated using a Roche LightCycler 480 2.0. Each 12.5 μL of qPCR reaction included 6.25 μL of GoTaq® qPCR Master Mix, 1.25 μl Dnase free Water, 2.5 μl of DNA template and 1.25 μL each of the 4.5 μM forward and reverse primers. Primers for the *Spiroplasma* dnaA gene were DnaA_F 5′-AAT GCT TGG ATC ATA ATT TAA AGA C-3′ and DnaA_R 5′-GTT TTG AAG AAA GAA ATG TTT CAA G-3′. Primers for the *A. pisum* Ef1a gene were Ef1a_F 5′-TAG CAG TTA CAT CAA GAA AAT CGG-3′ and Ef1a_R 5′-ATG TTG TCT CCA TTC CAT CCA G-3′. Cycling conditions are described in Table S2. Gene copy numbers were estimated with reference to a standard curve generated with serial dilutions of a synthetic standard. We did not standardize the overall DNA concentrations among samples because we were mainly interested in the *Spiroplasma* titers (number of *Spiroplasma* gene copies relative to aphid gene copies), and because the randomization of biological replicates safeguarded us against any unwanted biases. However, to improve the precision of the measurements, samples with a very high concentration were re-run after a dilution devised to yield an expected Cp around 20. For each sample, the number of gene copies per aphid individual was calculated from the average of triplicate qPCR reactions.

Because the format of the extraction kit (DNeasy 96 Blood & Tissue Kit [plate format] vs. DNeasy Blood & Tissue Kit [individual tubes]) had a strong effect on the estimated number of aphid gene copies and a minor effect on the estimated number of *Spiroplasma* gene copies (Fig. S1), we removed the estimated effect of the extraction kit using the function "removeBatchEffect" of the package "limma" (Smyth 2005, version 3.38.3) prior to further analyses. These corrected numbers of *Spiroplasma* and aphid gene copies per individual are indicated as #*Spiroplasma* dnaA and #aphid EF1a, respectively. The number of *Spiroplasma* gene copies per aphid gene copy is defined as #*Spiroplasma* dnaA/#aphid EF1a.

We fitted the "*Regiella*" and "*Spiroplasma*" models to each of the three variables #*Spiroplasma* dnaA, #aphid EF1a, and #*Spiroplasma* dnaA/#aphid EF1a. Since the uninfected sub-line was not included in this part of the study, the "*Regiella*" model did not contain the dummy variable "*S*" (i.e., all the investigated sub-lines carried *Spiroplasma*). The "*Regiella*" and "*Spiroplasma*" models additionally contained the aphid age (10 or 20 days) as a fixed effect as well as its interaction with the dummy variables "*R*" for the model "*Regiella*" and with the aphid sub-line for the model "*Spiroplasma*." The test procedure is the same a described for experiment 1, except that since there is no random effect in the "*Spiroplasma*" model, the main effects were tested using *F*-tests, and we additionally fitted a model separately for each aphid age to assess differences between sub-lines using Tukey's tests.

PHYLOGENETIC ANALYSES

We performed two analyses using the phylogeny of *Spiroplasma* strains inferred by Mathé-Hubert et al. (2018). This phylogeny (Fig. 1) showed that *Spiroplasma* of pea aphids are divided into at least three clades. The first analysis tested if the *Spiroplasma* induced phenotypes correlate with the phylogeny (phylogenetic inertia), which is expected if these phenotypes evolve slowly in comparison to the sequences used to discriminate *Spiroplasma* strains. Such phylogenetic inertia would mean that in pea aphids, different clades of *Spiroplasma* have different effects on their host. Then we tested if clade 3, which appears to have short branches in the phylogeny, has a lower rate of molecular evolution than the two other clades.

To test for phylogenetic inertia and to investigate the links among the *Spiroplasma* induced phenotypes, we characterized the variation in the effects of *Spiroplasma* strains on the phenotype of their host by the coefficients of the "*Spiroplasma*" models from the three experiments. These coefficients were used rather than the raw data because they represent the estimated effect of *Spiroplasma* after accounting for *Regiella*. A PCA was used to summarize this phenotypic variation. In this PCA, individuals (rows) are the *Spiroplasma* strains that are characterized by the coefficients of the "*Spiroplasma*" models on the different traits (i.e., one column per trait). These traits (columns) were weighted to ensure that the three experiments had the same weight whatever the number of traits we measured during the experiment. Since the intrinsic growth rate is a composite variable of other variables, it was included in the PCA as a supplementary variable: it was projected onto the PCA after the PCA was inferred. We tested for phylogenetic inertia on the first two PCA axes that jointly explained 57.03% of the phenotypic variation. Two measures of phylogenetic inertia are generally recommended, the lambda index and Abouheif's C_{mean} index (Münkemüller et al. 2012). For our phylogeny, the latter has more power (Fig. S2). Hence, we used C_{mean} to measure phylogenetic inertia and tested its significance by performing 10 000 randomizations using the package "phylosignal" (Keck et al. 2016).

For the *Spiroplasma* strains that share their host with *R. insecticola*, the coefficients used in the analysis describe the

Table 1. Analyses of parasitism by the parasitoid wasp *Aphidius ervi*.

Explained variable	Model		Wasp lines	Effect	df	F	p.value
Proportion mummified	**1**	"*Regiella*"	All	Wasp	2, 210	5.98	0.003**
				Spiro. (0/1)	1, 78.73	0.09	0.763
				Regi. (0/1)	1, 10	1.83	0.206
				Wasp × *Spiro*.	2, 210	0.44	0.646
				Wasp × *Regi*.	2, 210	0.18	0.836
	2	"*Spiroplasma*"	All	Wasp	2, 190	58.82	<0.001***
				Sub-line	12, 190	2.17	0.009**
				Wasp × Sub-line	24, 190	1.57	0.050.
	3	"*Spiroplasma*"	B	Sub-line	12, 60	4.91	<0.001***
	4	"*Spiroplasma*"	D	Sub-line	12, 60	0.73	0.72
	5	"*Spiroplasma*"	K	Sub-line	12, 60	1.25	0.273
Proportion emerged	**6**	"*Regiella*"	All	Wasp	2, 185.32	0.64	0.528
				Spiro. (0/1)	1, 160.60	0.28	0.594
				Regi. (0/1)	1, 10.53	3.08	0.108
				Wasp × *Spiro*.	2, 186.07	0.27	0.758
				Wasp × *Regi*.	2, 186.91	0.26	0.770
	7	"*Spiroplasma*"	All	Wasp	2, 164.64	5.63	0.004**
				Sub-line	12, 164.70	0.55	0.880
				Wasp × Sub-line	24, 164.65	0.87	0.638

Models 1–5 explain the proportion of the sets of 20 nymphs exposed to one wasp that were mummified. Models 6 and 7 explain the proportion of mummies from which a wasp emerged. Models 1 and 6 estimate the effect of *R. insecticola* and were used to build the offsets correcting for the presence of *R. insecticola* in the other models. Models 3–5 investigate the interaction between wasp line × aphid sub-line that is significant in model 2.

effect of the strain plus its potential interaction with *R. insecticola*. However, because strains with and without *R. insecticola* are similarly distributed in the phylogeny, potential interactions would only add noise to the analysis. This would decrease statistical power and thus should not create any false positives.

In the *Spiroplasma* phylogeny, clade 3 appears to have a lower rate of molecular evolution than clades 1 and 2. We used the local-clock permutation test developed by Lanfear (2010) to assess whether this difference was significant. This test is independent of the above mentioned experiments and only concerns the molecular phylogeny. It uses the ratio between the likelihood of two models that are fitted to the phylogeny and its underlying sequences (GenBank IDs MG288511 to MG288588). The first model assumes a strict clock, meaning that all strains are evolving equally fast, while in the second model (local clocks), strains of clade 3 are allowed to evolve at a different rate than other strains. The *P*-value is obtained by comparing the observed ratio between the likelihoods of the two models to the null distribution of this ratio, which is estimated by refitting the strict and the local clock models to 10,000 permutations of the sequences. This test has been shown to be more conservative than the usual likelihood ratio-test (Lanfear 2010). The local clocks model applied to the real data was also used to estimate the effect size of the difference of rates of molecular evolution.

Results

EXPERIMENT 1: EFFECT OF *SPIROPLASMA* ON *A. ervi* PARASITISM

The "*Regiella*" model detected highly significant variation among wasp lines in the proportion of aphids that were mummified (i.e., parasitized successfully), but no overall effects of the presence of either *Regiella* or *Spiroplasma* (Table 1). The "*Spiroplasma*" model also recovered the strong differences among wasp lines, with line B being the most and line D the least virulent line (Fig. 2), as well as significant variation among aphid sub-lines, also in interaction with the wasp line (Table 1). Separate analyses for each wasp line showed that this was mostly due to variation in susceptibility to the most virulent wasp line B (Table 1), for which the presence of *Spiroplasma* strains S227 and S385(+Ri) reduced parasitism significantly (Table S3). In the case of wasp line K, aphids infected with strain S161 were more likely to be successfully parasitized than the uninfected control sub-line (Table S3). Wasp lines also differed in proportion emerged, line K having the highest and line D the lowest emergence rate. However, this difference was detected by model "*Spiroplasma*" but not by model "*Regiella*," likely because of the higher complexity of the latter model.

EXPERIMENT 2: FITNESS COST OF *SPIROPLASMA*

The overall effect of the symbionts *Spiroplasma* and *R. insecticola* on the fitness of their host is summarized by the intrinsic growth

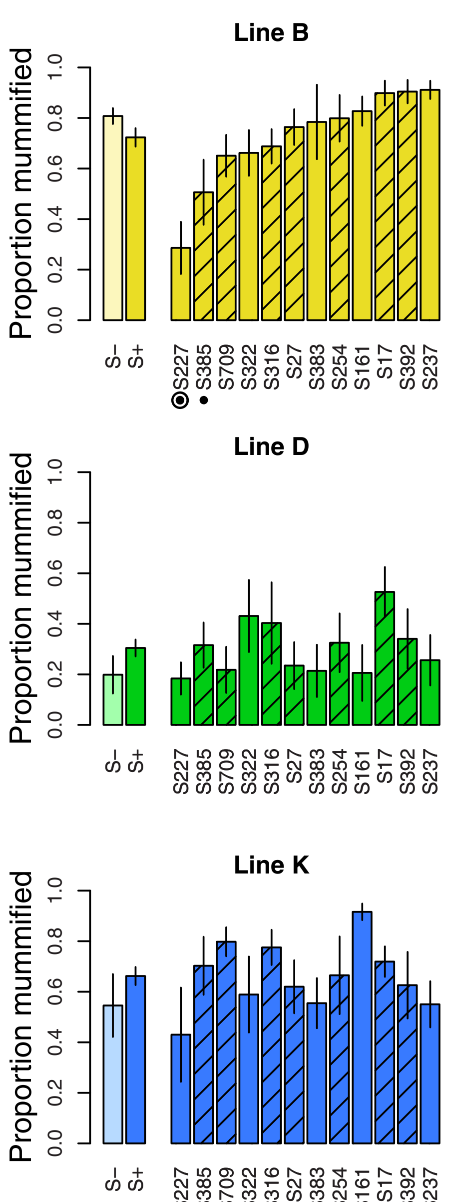


Figure 2. Mummification rates. Mean proportion of nymphs mummified (±S.E. indicated with error bars) for the three wasp lines and each aphid sub-line. On each panel, S– corresponds to the uninfected subline, and S+ to the mean of all *Spiroplasma* infected sub-lines. Error bars indicate the standard error. Sub-lines also containing *R. insecticola* are hatched. The significance of the comparisons between the *Spiroplasma* infected sub-lines and the uninfected control sub-lines performed from the 'Spiroplasma' models is indicated below the strains names (•: only significant before adjusting for multiple comparisons; ⊙: still significant after adjusting for multiple comparisons).

rate. We repeat the caveat that the interpretation of these effects hinges on strong assumptions, namely that the average effect of *Spiroplasma* strains that are alone is comparable to that of strains that are sharing their host with *R. insecticola*, and that there are no interactive effects of *Spiroplasma* and *R. insecticola* on aphid phenotypes. Under these—admittedly untested—assumptions, it appears that *Spiroplasma* reduced the intrinsic growth rate significantly while *R. insecticola* increased it or at least counteracted the negative effect of *Spiroplasma* (Table 2 and Fig. 3A). Correcting for the estimated effect of *R. insecticola*, the "*Spiroplasma*" model shows that all *Spiroplasma* strains except S322, S383, and S237 decreased the intrinsic growth rate significantly. This was still significant for more than half of the strains after correcting for multiple testing (Table S3).

Infection by *R. insecticola* did not affect aphid lifespan, but all *Spiroplasma*-infected sub-lines had shorter lifespans than the *Spiroplasma*-free sub-line, on average by about eight days (Table 2, Fig. 3B). Only the effect of strains S27 and S385(+Ri) on host survival was no longer significant after accounting for multiple testing (Table S3). *Spiroplasma* also reduced lifetime reproduction while *R. insecticola*—with the caveat mentioned above—appeared to increase it or at least to counteract the negative effect of *Spiroplasma* (Table 2 and Fig. 3C). Neither infection with *R. insecticola* nor infection with *Spiroplasma* had a significant overall effect on the mean reproductive age of the aphid host (Tables 2 and S3).

EXPERIMENT 3: VARIATION IN *SPIROPLASMA* DENSITY

Infection by *R. insecticola* did not have any detectable effect on #*Spiroplasma* dnaA, #aphid EF1a, or their ratio in either 10- or 20-day-old aphids (Table 3). The #aphid EF1a did not change significantly from age 10 to 20, but #*Spiroplasma* dnaA increased strongly (Table 3, Fig. 4B and C), on average by a factor of 4.86, which corresponds to an average doubling time of 4.38 days for *Spiroplasma*. Accordingly, the ratio of *Spiroplasma* to aphid gene copies increased as well and reached very high values (approximately 40–130) in 20-day-old aphids. There was substantial variation in the densities and growth achieved by different *Spiroplasma* strains, reflected in the highly significant sub-line and age × sub-line effects on #*Spiroplasma* dnaA (Table 3). This variation appeared to have a limited effect on aphid gene copy number, as the differences among sub-lines for #aphid EF1a were not statistically significant ($P = 0.08$, Table 3). *Spiroplasma* strain S227 was notable, however, because this sub-line showed very low #aphid EF1a in 10-day-old aphids, resulting in a high ratio of #*Spiroplasma* dnaA/#aphid EF1a (Fig. 4A). This is the sub-line that exhibited the lowest susceptibility to parasitoids but also high costs of infection by *Spiroplasma* (Figs. 2 and 3).

PHYLOGENETIC SIGNAL IN *SPIROPLASMA* PHENOTYPES AND RATE OF MOLECULAR EVOLUTION

The first two axes of the PCA that were tested for a phylogenetic signal summarized 57.03% of the phenotypic variation in the 12 *Spiroplasma*-infected pea aphid sub-lines. The first axis

Table 2. Analyses of the fitness costs of *Spiroplasma*.

Explained variable	Model		Effect	df	F (or χ^2)#	p-value
Lifetime reproduction	**1**	"*Regiella*"	*Spiro*. (0/1)	1, 41.86	9.27	0.004**
			Regi. (0/1)	1, 10.06	6.61	0.028*
	2	"Spiroplasma"	Sub-line	12, 90.21	5.46	<0.001***
Lifespan	**3**	"*Regiella*"	*Spiro*. (0/1)	1, 173.99	11.46	<0.001***
			Regi. (0/1)	1, 173.99	0.73	0.392
	4	"Spiroplasma"	Sub-line	12, 172.99	95.61	<0.001***
Mean reproductive age	**5**	"*Regiella*"	*Spiro*. (0/1)	1, 81.66	0.18	0.673
			Regi. (0/1)	1, 10.07	1.65	0.227
	6	"Spiroplasma"	Sub-line	12, 141.42	4.58	<0.001***
Intrinsic growth rate	**7**	"*Regiella*"	*Spiro*. (0/1)	1, 51.57	8.84	0.004**
			Regi. (0/1)	1, 10.01	10.61	0.009**
	8	"Spiroplasma"	Sub-line	12, 85.97	4.19	<0.001***

Models 1, 3, 5, and 7 estimate the effect of *R. insecticola* on four variables related to fitness. They were used to build the offsets correcting for the presence of *R. insecticola* in the other models estimating the effect of each *Spiroplasma* strain (models 2, 4, 6, and 8).

#For lifespan, we used a Cox model, for which fixed effect were tested with LRT. In this case, we report the χ^2 statistics.

Table 3. Analyses of the *qPCR* estimates of the number of gene copies in 10 and 20 days old aphids.

Explained variable	Model		Aphid age	Effect	*df*	*F*	*P*-value
#*Spiro*. dnaA/#aphid EF1a	**1**	"*Regiella*"	Both	*Regi*. (0/1)	1, 10.01	121.34	0.765
				Age	1, 106.05	0.25	<0.001***
				Age × *Regi*.	1, 106.20	2.12	0.148
	2	"*Spiroplasma*"	Both	Age	1, 96	314.95	<0.001***
				Sub-line	11, 96	8.07	<0.001***
				Age × Sub-line	11, 96	1.96	0.041*
	3	"*Spiroplasma*"	10 days	Sub-line	11, 49	4.50	<0.001***
	4	"*Spiroplasma*"	20 days	Sub-line	11, 47	5.38	<0.001***
#*Spiro*. dnaA	**5**	"*Regiella*"	Both	*Regi*. (0/1)	1, 10.01	0.00	0.993
				Age	1, 106.04	262.80	<0.001***
				Age × *Regi*.	1, 106.14	0.00	0.985
	6	"*Spiroplasma*"	Both	Age	1, 96	728.20	<0.001***
				Sub-line	11, 96	12.31	<0.001***
				Age × Sub-line	11, 96	2.63	0.006**
	7	"*Spiroplasma*"	10 days	Sub-line	11, 49	5.66	<0.001***
	8	"*Spiroplasma*"	20 days	Sub-line	11, 47	10.02	<0.001***
#aphid EF1a	**9**	"*Regiella*"	Both	*Regi*. (0/1)	1, 10.05	1.31	0.278
				Age	1, 106.22	0.25	0.618
				Age × *Regi*.	1, 106.83	4.00	0.048*
	10	"*Spiroplasma*"	Both	Age	1, 96	0.68	0.412
				Sub-line	11, 96	1.70	0.084.
				Age × Sub-line	11, 96	1.16	0.325

Models 1, 5, and 9 estimate the effect of *R. insecticola* on #*Spiroplasma* dnaA/#aphid EF1a, #*Spiroplasma* dnaA, and #aphid EF1a. They were used to build the offsets correcting for the presence of *R. insecticola* in the other models estimating the effect of each *Spiroplasma* strain (models 2, 6, and 10). When there was a significant interaction between aphid age and sub-line, separate models were fitted for each age group to investigate the interaction (models 3, 4, 7, and 8).

mainly summarized the negative effect that *Spiroplasma* strains with a high density had on the lifespan of their host (Fig. 5A). This negative effect on lifespan had little effect on the aphids' intrinsic growth rate because this first axis has only a low correlation with the lifetime reproduction and a negative correlation with the mean reproductive age (i.e., short-lived aphids produced offspring earlier in life). The second axis encompasses variation related to aphid health and suitability for parasitoids. This axis was positively correlated to #aphid EF1a, the aphid growth rate, the lifetime reproduction, and negatively correlated to the mean

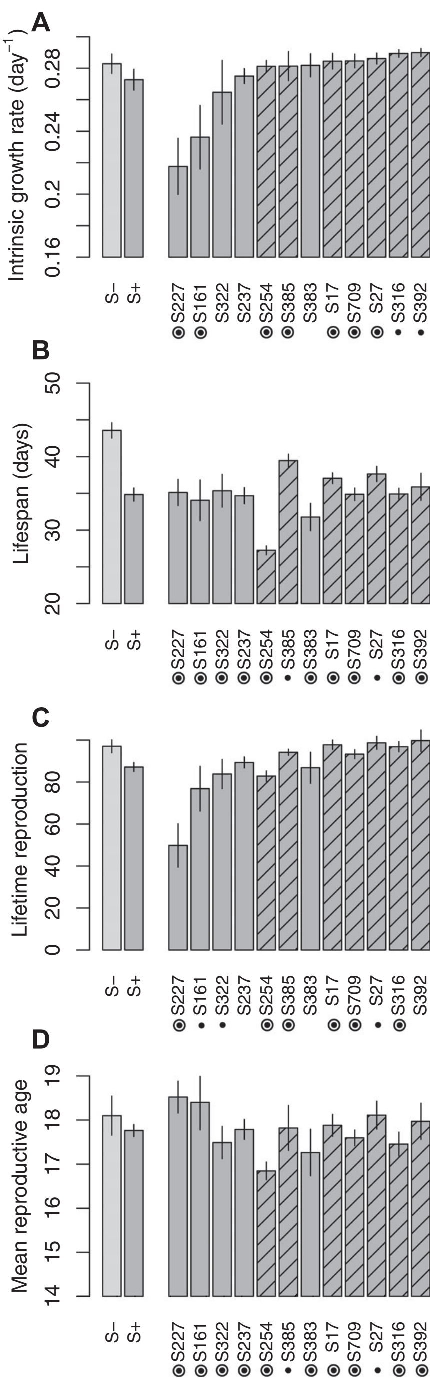


Figure 3. *Spiroplasma* **effects on aphid fitness. Bar plots depicting the average intrinsic growth rate (A), lifespan (B), lifetime reproduction (C), and mean reproductive age (D) for all aphid sub-lines. On each panel, S– corresponds to the uninfected sub-line, and S+ to the mean of all** *Spiroplasma* **infected sub-lines. Error bars indicate the standard error. Sub-lines also containing** ***R. insecticola*** **are hatched. The significance of the comparisons between the** *Spiroplasma* **infected sub-lines and the uninfected control sub-lines performed from the "*Spiroplasma*" models is indicated below the strains' names (•: only significant before adjusting for multiple comparisons; ⊙: still significant after adjusting for multiple comparisons).**

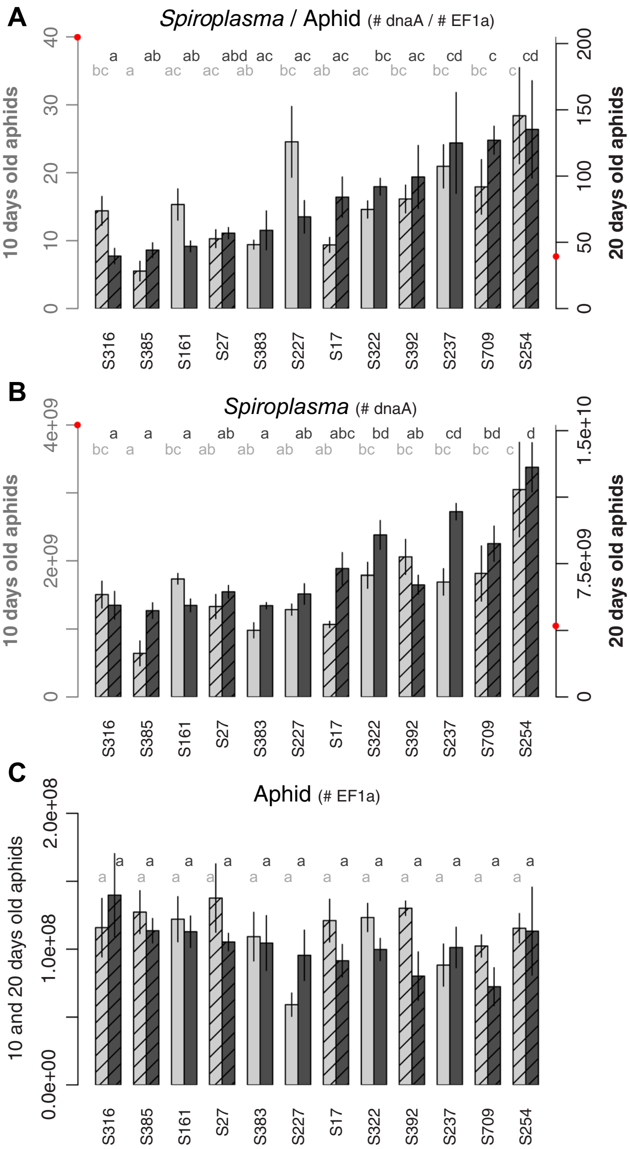


Figure 4. *Spiroplasma* **density. The number of** *Spiroplasma* **gene copies #per aphid gene copy (#*Spiroplasma* dnaA/# aphid EF1a) as well as the raw numbers of** *Spiroplasma* **and aphid gene copies per aphid individual (#** *Spiroplasma* **dnaA and # aphid EF1a) are shown on panels (A–C). Because #** *Spiroplasma* **dnaA is much higher in 20 days old aphids (dark grey) than in 10 days old aphids (light grey), panels (A) and (B) have two** *y***-axes with different scales. To help the comparison, red dots indicate the same values on the left and right axes. Different letters above bars indicate significant pairwise differences in Tukey–HSD tests. Error bars indicate the standard errors.**

reproductive age. Sub-lines with a higher score on this axis (i.e., more fecund sub-lines) also showed higher rates of mummification by parasitoids and parasitoid emergence (Fig. 5B). The variation in the reproductive fitness of the sublines was not a function of *Spiroplasma* titers, as the variation in #*Spiroplasma* dnaA was only weakly correlated with this axis.

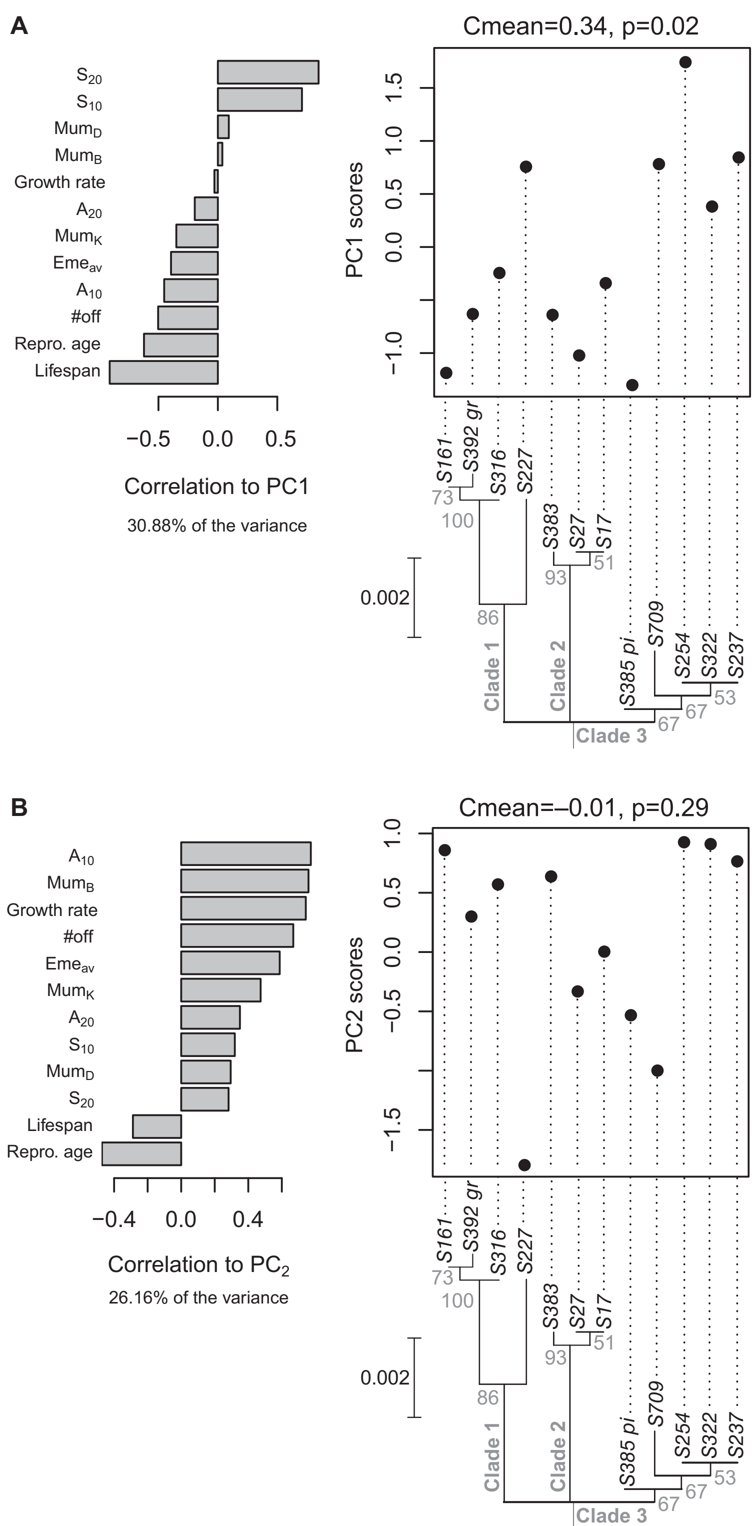


Figure 5. Phylogenetic signal in the *Spiroplasma* induced phenotypes. (A and B) Phylogenetic signal in principal components (PC) 1 and 2, summarizing the phenotypic variation induced by the different *Spiroplasma* strains. The PCA was done on the coefficients of the models fitted to the phenotypic traits of the aphids. On both panels, the horizontal bar plot indicates the correlation between the model coefficients that are the variables of the PCA and the corresponding PCA axis. This informs about the phenotypes of the aphids that are summarized by the PCA axis. Variable names were shortened: S and A refer to #*Spiroplasma* dnaA and #aphid EF1a in 10 and 20 days old aphids; Mum and Eme refer to the mummification and emergences rates, respectively, in wasps B, D, and K and averaged ($_{av}$) over the three wasp lines; #off refers to lifetime reproduction. The plots on the right side give the score of each *Spiroplasma* strain on the PCA axes and its position in the phylogeny. The scale bar indicates the substitution rate.

These two axes were used to investigate the correlation between the *Spiroplasma*-induced phenotypic variation and the *Spiroplasma* phylogeny using the Abouheif's C_{mean} statistic. Only the first axis was significantly correlated (PC1: $C_{mean} = 0.34$, $P = 0.02$; PC2: $C_{mean} = -0.01$, $P = 0.29$), with most strains of clade 3 having a high score on the first axis (Fig. 5A).

The local clock model estimated that the sequences of clade 3 are evolving 5.6 times more slowly than those of clades 1 and 2. The local-clock permutation test revealed that this difference was marginally significant ($P = 0.043$).

Discussion

In the absence of reproductive manipulation or frequent horizontal spread, heritable endosymbionts must provide a net fitness benefit to persist in host populations (Oliver et al. 2014). We investigated protection against the parasitoid wasp *A. ervi* as a potential benefit provided by 12 different strains of *Spiroplasma* in pea aphids, and we estimated their costs to the host in terms of life-history traits.

Evidence for protection was very limited and restricted to one of the three lines of *A. ervi* we used. Only *Spiroplasma* strains S227 and S385 reduced parasitism by the most virulent wasp line B significantly. In the case of S227, however, this was associated with very low reproductive fitness of the aphids in the absence of parasitoids, suggesting that S227-infected aphids were generally of poor health. On the other hand, when the aphids were exposed to wasp line K, one strain of *Spiroplasma* (S161) even seemed to represent a significant liability and made aphids more susceptible to parasitism. The effects of some *Spiroplasma* strains tended to be unequal across the three parasitoid lines, which resulted in a near-significant genotype-by-genotype interaction (Table 1). In principle, such interactions could contribute to the maintenance of strain diversity in parasites as well as symbionts (Kwiatkowski and Vorburger 2012; Ford et al. 2017; Vorburger and Perlman 2018), although their importance is questionable here, because the majority of *Spiroplasma* strains had no detectable effects on parasitism. We do not know why the three wasp lines varied so strongly in their parasitism efficacy. The conspicuously low success of line D could be related to the long time it has been reared in our laboratory at relatively small population size, which might have resulted in negative effects of inbreeding. The difference between the two commercially available lines may be related to their long-term rearing conditions in the production and/or their genetic background. Genetic variation in parasitism success is commonplace in parasitoids (Kraaijeveld and Godfray 1999; Colinet et al. 2010; Sandrock et al. 2010) and likely related to variation in the cocktail of virulence factors parasitoids employ. For example, parasitoid wasp venom is a major source of virulence factors and generally shows a high level of intraspecific variation (Colinet et al. 2013; Mathé-Hubert et al. 2015), also in *A. ervi* (Colinet et al. 2014). Interactions between parasitoid virulence factors and *Spiroplasma* in the aphid hosts could potentially explain the somewhat uneven effects of the different *Spiroplasma* strains in the three parasitoid treatments.

Even though we find little evidence for protection against *A. ervi* in the present study, it should be added that *Spiroplasma* may still reduce the risk of parasitism indirectly via a plant-mediated effect, because *A. ervi* is more attracted to volatiles from plants infested by *Spiroplasma*-free aphids than from plants with *Spiroplasma*-infected aphids, as recently shown by Frago et al. (2017). Such an effect would have been missed by our non-choice assays.

Due to an unfortunate handling error in the preparation of our experimental lines, about half of the *Spiroplasma* strains shared their hosts with a coinfection of *R. insecticola*. However, the presence of *R. insecticola* did not have any detectable effects on susceptibility to *A. ervi*. This outcome is consistent with earlier studies that tested the same strain of *R. insecticola* deliberately and reported no significant effects on parasitism by *A. ervi* (Oliver et al. 2003; Hansen et al. 2012).

In contrast to the potential benefits we investigated, the costs of infection with *Spiroplasma* were rather clear. All strains curtailed aphid lifespan significantly, on average by more than 8 days. A reduction of host lifespan is also characteristic of *Spiroplasma* infection in *Drosophila melanogaster* (Herren et al. 2014). Because old aphids are less fecund than young adults (e.g., Zeng et al. 1993; Vorburger and Ramsauer 2008), and because offspring produced early in life are more important for a clone's growth rate than offspring produced late (Lenski and Service 1982), the strong negative effect on lifespan did not translate into equally strong effects on lifetime reproduction and the intrinsic rate of increase (Fig. 3). Nevertheless, two of the five sub-lines infected only by *Spiroplasma* showed significantly lower intrinsic rates of increase than the uninfected control. The sub-lines in which *Spiroplasma* co-occurred with *R. insecticola* showed similar trait values to the uninfected sub-line. This would suggest that the presence of *R. insecticola* counterbalanced the costs imposed by *Spiroplasma*. The "*Regiella*" models indeed showed a positive overall effect of *R. insecticola* on lifetime reproduction as well as the intrinsic rate of increase. This interpretation of the results is, however, conditional on the validity of the assumption that *Spiroplasma* strains associated with *R. insecticola* have the same average effect as those that are not. A positive effect of *R. insecticola* on host fitness has also been reported by Tsuchida et al. (2004) for pea aphids feeding on clover, but this does not seem to be a general property of this symbiont (Leonardo 2004; Ferrari et al. 2007), and other studies reported negative fitness effects of this symbiont, for example, in the grain aphid, *Sitobion avenae* (Wang et al. 2016; Luo et al. 2017). Nonetheless, it has been observed before that one heritable

endosymbiont can compensate the costs imposed by another. Doremus and Oliver (2017) found that the large costs associated with the possession of X-type endosymbionts in pea aphids were ameliorated by coinfection with *H. defensa*. When the influence of *R. insecticola* was corrected for statistically in the present data, the majority of *Spiroplasma* strains were inferred to reduce lifetime reproduction and intrinsic rate of increase significantly (Table S3). Thus we conclude that under laboratory conditions and in the absence of any other selective forces, infection with *Spiroplasma* generally has a negative effect on pea aphid reproductive fitness.

To some extent, the *Spiroplasma*-induced fitness costs were related to the symbionts' densities in the host. The *Spiroplasma* titers in pea aphids increased strongly from the age of 10 days to the age of 20 days, suggesting that the host has limited control over the proliferation of *Spiroplasma*. This is also observed in *D. melanogaster*, and it may be related to the fact that cell wall-less bacteria like *Spiroplasma* can escape the attention of the insect immune system (Herren and Lemaitre 2011; Herren et al. 2014). However, not all strains were equally prolific. *Spiroplasma* titers varied substantially among aphid sub-lines, and there was no indication that they were influenced by coinfecting *R. insecticola*. The links among the estimated effects of the different *Spiroplasma* strains on the various traits we measured was investigated with a PCA on the coefficients estimated by the models analyzing these traits. In this PCA, the first PC was chiefly associated with high *Spiroplasma* densities and short aphid lifespan, providing correlative evidence that higher *Spiroplasma* titers are more harmful to the host. Interestingly, there was a weak but significant phylogenetic signal in the variation along this axis (Fig. 5). This was mostly because strains from clade 3 showed higher scores for PC1 on average, i.e., these strains achieved higher densities and tended to be associated with shorter host lifespans. High *Spiroplasma* densities have also been shown to curtail host lifespan in flies (Herren and Lemaitre 2011). Clade 3 also exhibited a lower rate of molecular evolution than the other two clades, and it is tempting to speculate about a causal link with the seemingly more parasitic lifestyle of these *Spiroplasma* strains. Endosymbiotic bacteria generally exhibit increased rates of sequence evolution than their free-living relatives, which is attributed to the lower effective population size that comes with maternal transmission and the associated bottlenecks between host generations (Moran 1996; Woolfit and Bromham 2003; Boscaro et al. 2013). Long-term vertical transmission is also expected to reduce the costs that symbionts impose on their hosts. Endosymbionts are thus a good model of how organisms can move along the parasitism–mutualism continuum (Ewald 1987; King 2019). It might seem that *Spiroplasma* strains from clade 3 occupy a space further toward the parasitic end of this continuum than the other clades. Whether this reflects a shorter association with the host, which would be consistent with the slower rate of molecular evolution, or whether other selective forces have created this situation, is currently unknown. In this context it could be relevant that the different clades of *Spiroplasma* tend to be associated with different communities of co-infecting symbionts in natural populations of pea aphids. For example, clade 3 *Spiroplasma* are positively associated in the field with the X-type symbiont and negatively with *H. defensa*, while those from clade 2 tend to be positively associated with *Rickettsia*, and this seems to be unrelated to the host plants from which pea aphids were collected (Mathé-Hubert et al. 2018). Regular coinfections with other bacteria certainly have the potential to affect the evolution of endosymbionts and possibly their virulence (Vorburger and Perlman 2018). Interactions with other species of endosymbiotic bacteria thus clearly deserve attention to better understand *Spiroplasma*'s influence on host ecology and evolution.

In conclusion, our experiment showed that infections with various strains of the heritable endosymbiont *Spiroplasma* are rather costly to their pea aphid hosts, and that protection against *A. ervi* is unlikely to compensate for these costs. We tested for protection against *A. ervi* because it is the pea aphid's most common parasitoid, but multiple parasitoids include pea aphids in their host range and we cannot exclude that *Spiroplasma* may be protective against other species. Already demonstrated is a protective effect of certain strains of *Spiroplasma*, including strain S161 used here, against the entomopathogenic fungus *Pandora neoaphidis* (Łukasik et al. 2013), but also this is not a general property of all *Spiroplasma* found in pea aphids. Once a symbiont has evolved maternal transmission, it is under strong selection to keep its host alive until reproduction. This can be achieved via protection against natural enemies or by providing other ecological benefits, for example, by increasing tolerance to abiotic stressors (Oliver et al. 2010). The specific mechanisms may well vary among different strains of the same symbiont species, and with the high diversity of strains structured into at least three clades, *Spiroplasma* of pea aphids is an attractive model to investigate this variation further.

AUTHOR CONTRIBUTIONS

H.K., P.G., C.V., and H.M.H. performed the experiments; P.G., H.M.H., and H.K. carried out the molecular analysis; H.M.H. was responsible for the data analysis; H.M.H., H.K., and C.V. wrote the paper.

ACKNOWLEDGMENTS

The authors are very grateful to Paula Rodriguez for her help with insect rearing. The authors also thank Ailsa McLean for sharing several *Spiroplasma*-infected aphid lines. The handling editor and the reviewers provided very helpful comments to improve the manuscript. This work was supported by a Sinergia grant from the Swiss National Science Foundation (grant nr. CRSII3_154396 to C.V.).

DATA ARCHIVING

The DNA sequences used in this study are available in Genbank (accession numbers: MG288511 to MG288588). The main dataset is deposited in the Dryad Digital Repository: https://datadryad.org/stash/dataset/doi:10.5061/dryad.q0r84n4

LITERATURE CITED

Asplen, M. K., N. Bano, C. M. Brady, N. Desneux, K. R. Hopper, C. Malouines, K. M. Oliver, J. A. White, and G. E. Heimpel. 2014. Specialisation of bacterial endosymbionts that protect aphids from parasitoids. Ecol. Entomol. 39:736–739.

Ballinger, M. J., and S. J. Perlman. 2018. The defensive spiroplasma. Curr. Opin. Insect Sci. 32:1–6.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8.

Bian, G., Y. Xu, P. Lu, Y. Xie, and Z. Xi. 2010. The endosymbiotic bacterium *Wolbachia* induces resistance to dengue virus in *Aedes aegypti*. PLoS One 6:e1000833.

Birch, L. C. 1948. The intrinsic rate of natural increase of an insect population. J. Anim. Ecol. 17:15–26.

Boscaro, V., M. Felletti, C. Vannini, M. S. Ackerman, P. S. G. Chain, S. Malfatti, L. M. Vergez, M. Shin, T. G. Doak, M. Lynch, et al. 2013. *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. Proc. Natl. Acad. Sci. USA 110:18590–5.

Bové, J. M., J. Renaudin, C. Saillard, X. Foissac, and M. Garnier. 2003. *Spiroplasma citri*, a plant pathogenic mollicute: relationships with its two hosts, the plant and the leafhopper vector. Annu. Rev. Phytopathol. 41:483–500.

Caillaud, M. C., M. Boutin, C. Braendle, and J.-C. Simon. 2002. A sex-linked locus controls wing polymorphism in males of the pea aphid, *Acyrthosiphon pisum* (Harris). Heredity (Edinb). 89:346–352.

Cayetano, L., and C. Vorburger. 2013. Genotype-by-genotype specificity remains robust to average temperature variation in an aphid/endosymbiont/parasitoid system. J. Evol. Biol. 26:1603–10.

Cayetano, L., and C. Vorburger. 2014. Symbiont-conferred protection against Hymenopteran parasitoids in aphids: how general is it? Ecol. Entomol. 40:1–9.

Chen, X., S. Li, and S. Aksoy. 1999. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J. Mol. Evol. 48:49–58.

Chiel, E., Y. Gottlieb, N. Katzir, M. Inbar, and M. Ghanim. 2007. Biotype-dependent secondary symbiont communities in sympatric populations of *Bemisia tabaci*. Bull. Entomol. Res. 97:407–413.

Clark, T. B., R. F. Whitcomb, J. G. Tully, C. Mouches, C. Saillard, J. M. Bové, H. Wroblewski, P. Carle, D. L. Rose, R. B. Henegar, et al. 1985. *Spiroplasma melliferum*, a new species from the Honeybee (*Apis mellifera*). Int. J. Syst. Bacteriol. 35:296–308.

Colinet, D., C. Anselme, E. Deleury, D. Mancini, J. Poulain, C. Azéma-Dossat, M. Belghazi, S. Tares, F. Pennacchio, M. Poirié, et al. 2014. Identification of the main venom protein components of *Aphidius ervi*, a parasitoid wasp of the aphid model *Acyrthosiphon pisum*. BMC Genomics 15: 1–14.

Colinet, D., H. Mathé-Hubert, R. Allemand, J. L. J. L. Gatti, and M. Poirié. 2013. Variability of venom components in immune suppressive parasitoid wasps: from a phylogenetic to a population approach. J. Insect Physiol. 59:205–12.

Colinet, D., A. Schmitz, D. Cazes, and J. Gatti. 2010. The origin of intraspecific variation of virulence in an eukaryotic immune suppressive parasite. PLoS Pathog. 6:1–11.

Wang, D., X. Shi, P. Dai, D. Liu, X. Dai, Z. Shang, Z. Ge, and X. Meng. 2016. Comparison of fitness traits and their plasticity on multiple plants for *Sitobion avenae* infected and cured of a secondary endosymbiont. Sci. Rep. 6:23177.

Doremus, M. R., and K. M. Oliver. 2017. Aphid heritable symbiont exploits defensive mutualism. Appl. Environ. Microbiol. 83:1–15.

Douglas, A. E. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria buchnera. Annu. Rev. Entomol. 43:17–37.

Duron, O., D. Bouchon, S. Boutin, L. Bellamy, L. Zhou, J. Engelstadter, and G. D. Hurst. 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. BMC Biol. 6:27.

Ewald, P. W. 1987. Transmission modes and evolution of the parasitism-mutualism continuum. Ann. N. Y. Acad. Sci. 503:295–306.

Ferrari, J., C. L. Scarborough, and H. C. J. Godfray. 2007. Genetic variation in the effect of a facultative symbiont on host-plant use by pea aphids. Oecologia 153:323–329.

Ferrari, J., J. a West, S. Via, and H. C. J. Godfray. 2012. Population genetic structure and secondary symbionts in host-associated populations of the pea aphid complex. Evolution (N. Y). 66:375–390.

Ford, S. A., D. Williams, S. Paterson, and K. C. King. 2017. Co-evolutionary dynamics between a defensive microbe and a pathogen driven by fluctuating selection. Mol. Ecol. 26:1778–1789.

Frago, E., M. Mala, B. T. Weldegergis, C. Yang, A. McLean, H. C. J. Godfray, R. Gols, and M. Dicke. 2017. Symbionts protect aphids from parasitic wasps by attenuating herbivore-induced plant volatiles. Nat. Commun. 8:1–9.

Gasparich, G. E., R. F. Whitcomb, D. Dodge, F. E. French, J. Glass, and D. L. Williamson. 2004. The genus *Spiroplasma* and its non-helical descendants: Phylogenetic classification, correlation with phenotype and roots of the *Mycoplasma mycoides* clade. Int. J. Syst. Evol. Microbiol. 54:893–918.

Halekoh, U., and S. Højsgaard. 2014. A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models - the R package pbkrtest. J. Stat. Softw. 59:1–32.

Hansen, A. K., C. Vorburger, and N. A. Moran. 2012. Genomic basis of endosymbiont-conferred protection against an insect parasitoid. Genome Res. 22:106–114.

Harrell, F. E. 2017. rms: Regression modeling strategies. R package version 5:1–1.

Harrison, X. A. 2015. A comparison of observation-level random effect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. PeerJ 3:e1114.

Heath, K. D., and J. R. Stinchcombe. 2014. Explaining mutualism variation: a new evolutionary paradox? Evolution 68:309–317.

Hedges, L. M., J. C. Brownlie, S. L. O'Neill, and K. N. Johnson. 2008. *Wolbachia* and virus protection in insects. Science 322:702.

Henry, L. M., J. Peccoud, J. C. Simon, J. D. Hadfield, M. J. C. Maiden, J. Ferrari, and H. C. J. Godfray. 2013. Horizontally transmitted symbionts and host colonization of ecological niches. Curr. Biol. 23:1713–1717.

Herren, J. K., and B. Lemaitre. 2011. *Spiroplasma* and host immunity: activation of humoral immune responses increases endosymbiont load and susceptibility to certain Gram-negative bacterial pathogens in *Drosophila melanogaster*. Cell. Microbiol. 13:1385–1396.

Herren, J. K., J. C. Paredes, F. Schüpfer, K. Arafah, P. Bulet, B. Lemaitre, L. Epfl, P. B. Archamps, and S. Julien. 2014. Insect endosymbiont proliferation is limited by lipid availability. Elife 3:e02964.

Hilgenboecker, K., P. Hammerstein, P. Schlattmann, A. Telschow, and J. H. Werren. 2008. How many species are infected with *Wolbachia*? A statistical analysis of current data. FEMS Microbiol. Lett. 281:215–20.

Hothorn, T., F. Bretz, and P. Westfall. 2008. Simultaneous inference in general parametric models. Biometric. J. 50:346–363.

Hutchinson, M. K., and M. C. Holtman. 2005. Focus on research methods analysis of count data using Poisson regression*. Res. Nurs. Health 28:408–418.

Jaenike, J., R. Unckless, and S. Cockburn. 2010. Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. Science (80-.). 329:212–215.

Keck, F., F. Rimet, A. Bouchez, and A. Franc. 2016. phylosignal: an R package to measure, test, and explore the phylogenetic signal. Ecol. Evol. 6:2774–2780.

King, K. C. 2019. Defensive symbionts. Curr. Biol. 29:R78–R80.

Kraaijeveld, A. R., and H. C. J. Godfray. 1999. Geographic patterns in the evolution of resistance and virulence in drosophila and its parasitoids. Am. Nat. 153:61–74.

Kwiatkowski, M., J. Engelstädter, and C. Vorburger. 2012. On genetic specificity in symbiont-mediated host-parasite coevolution. PLoS Comput. Biol. 8:e1002633.

Kwiatkowski, M., and C. Vorburger. 2012. Modeling the ecology of symbiont-mediated protection against parasites. Am. Nat. 179:595–605.

Lanfear, R. 2010. The local-clock permutation test: a simple test to compare rates of molecular evolution on phylogenetic trees. Evolution 65:606–611.

Lenski, R. E., and P. M. Service. 1982. The statistical analysis of population growth rates calculated from schedules of survivorship and fecunidity. Ecology 63:655–662.

Leonardo, T. E. 2004. Removal of a specialization-associated symbiont does not affect aphid fitness. Ecol. Lett. 7:461–468.

Leybourne, D. J., J. I. Bos, T. A. Valentine, and A. J. Karley. 2018. The price of protection: a defensive endosymbiont impairs nymph growth in the bird cherry-oat aphid, *Rhopalosiphum padi*. Insect Science.

Łukasik, P., H. Guo, M. Van Asch, J. Ferrari, and H. C. J. Godfray. 2013. Protection against a fungal pathogen conferred by the aphid facultative endosymbionts *Rickettsia* and *Spiroplasma* is expressed in multiple host genotypes and species and is not influenced by co-infection with another symbiont. J. Evol. Biol. 26:2654–2661.

Luo, C., K. Luo, L. Meng, B. Wan, H. Zhao, and Z. Hu. 2017. Ecological impact of a secondary bacterial symbiont on the clones of *Sitobion avenae* (Fabricius) (Hemiptera: Aphididae). Nat. Publ. Gr. 7:1–8.

Martinez, A. J., K. L. Kim, J. P. Harmon, and K. M. Oliver. 2016. Specificity of multi-modal aphid defenses against two rival parasitoids. PLoS One 11:e0154670.

Mathé-Hubert, H., J.-L. Gatti, D. Colinet, M. Poirié, and T. Malausa. 2015. Statistical analysis of the individual variability of 1D protein profiles as a tool in ecology: an application to parasitoid venom. Mol. Ecol. Resour. 15:1120–1132.

Mathé-Hubert, H., H. Kaech, C. Hertaeg, and C. Vorburger. 2018. Non-random associations of maternally transmitted symbionts in insects: The roles of drift versus co-transmission and selection. bioRxiv 1–30. https://doi.org/10.1101/364653.

McLean, A. H. C. 2019. Cascading effects of defensive endosymbionts. Curr. Opin. Insect Sci. 32:42–46.

Mclean, A. H. C., and H. C. J. Godfray. 2015. Evidence for specificity in symbiont- conferred protection against parasitoids. Proc. R. Soc. B. 282:1–8.

McLean, A. H. C., and H. C. J. Godfray. 2017. The outcome of competition between two parasitoid species is influenced by a facultative symbiont of their aphid host. Funct. Ecol. 31:927–933.

McLean, A. H. C., M. van Asch, J. Ferrari, and H. C. J. Godfray. 2011. Effects of bacterial secondary symbionts on host plant use in pea aphids. Proc. Biol. Sci. 278:760–766.

Moran, N. A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA 93:2873–2878.

Moran, N. a, J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. 42:165–190.

Moran, N., and J. Russell. 2005. Evolutionary relationships of three new species of enterobacteriaceae living as symbionts of aphids and other insects. Appl. Environ. Microbiol. 71:3302–3310.

Münkemüller, T., S. Lavergne, B. Bzeznik, S. Dray, T. Jombart, K. Schiffers, and W. Thuiller. 2012. How to measure and test phylogenetic signal. Methods Ecol. Evol. 3:743–756.

Oliver, K. M., J. Campos, N. a Moran, and M. S. Hunter. 2008. Population dynamics of defensive symbionts in aphids. Proc. Biol. Sci. 275:293–299.

Oliver, K. M., P. H. Degnan, G. R. Burke, and N. a Moran. 2010. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. Annu. Rev. Entomol. 55:247–266.

Oliver, K. M., and N. A. Moran. 2009. Defensive symbionts in aphids and other insects. In: Defensive mutualism in microbial symbiosis, eds., White J. F. and Torres M. S. Taylor & Francis, Abingdon, U.K., pp. 129–147.

Oliver, K. M., J. A. Russell, N. A. Moran, and M. S. Hunter. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc. Natl. Acad. Sci. USA. 100:1803–7.

Oliver, K. M., A. H. Smith, and J. A. Russell. 2014. Defensive symbiosis in the real world-advancing ecological studies of heritable, protective bacteria in aphids and beyond. Funct. Ecol. 28:341–355.

Paredes, J. C., J. K. Herren, F. Schüpfer, and B. Lemaitre. 2016. The role of lipid competition for endosymbiont-mediated protection against parasitoid wasps in Drosophila. MBio 7:e01006–16.

Parker, B. J., J. Hrcek, A. H. C. McLean, and H. C. J. Godfray. 2017. Genotype specificity among hosts, pathogens, and beneficial microbes influences the strength of symbiont-mediated protection. Evolution 71:1222–1231.

Raychoudhury, R., L. Baldo, D. C. S. G. Oliveira, and J. H. Werren. 2009. Modes of acquisition of *Wolbachia*: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. Evolution 63:165–183.

Russell, J. a., S. Weldon, A. H. Smith, K. L. Kim, Y. Hu, P. Łukasik, S. Doll, I. Anastopoulos, M. Novin, and K. M. Oliver. 2013. Uncovering symbiont-driven genetic diversity across North American pea aphids. Mol. Ecol. 22:2045–2059.

Sandrock, C., A. Gouskov, and C. Vorburger. 2010. Ample genetic variation but no evidence for genotype specificity in an all-parthenogenetic host-parasitoid interaction. J. Evol. Biol. 23:578–585.

Scarborough, C. L., J. Ferrari, and H. C. J. Godfray. 2005. Aphid protected from pathogen by endosymbiont. Science 310:1781.

Schmid, M., R. Sieber, Y.-S. S. Zimmermann, and C. Vorburger. 2012. Development, specificity and sublethal effects of symbiont-conferred resistance to parasitoids in aphids. Funct. Ecol. 26:207–215.

Simon, J. C., S. Boutin, T. Tsuchida, R. Koga, J. F. Gallic, A. Frantz, Y. Outreman, and T. Fukatsu. 2011. Facultative symbiont infections affect aphid reproduction. PLoS One 6:e21831.

Smyth, G. K. 2005. Limma: linear models for microarray data. In: Bioinformatics and computational biology solutions using R and bioconductor, eds., Gentleman R., Carey V., Huber W., Dudoit S., and Irizarry R. Springer, New York, NY, pp. 397–420.

Teixeira, L., A. Ferreira, and M. Ashburner. 2008. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biol. 6:e1000002.

The International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. 8:e1000313.

Therneau, T. M. 2015a. coxme: mixed effects cox models. 1–14.

Therneau, T. M. 2015b. A package for survival analysis in S. version 2.38.

Tsuchida, T., R. Koga, and T. Fukatsu. 2004. Host plant specialization governed by facultative symbiont. Science 303:1989.

Vorburger, C. 2014. The evolutionary ecology of symbiont-conferred resistance to parasitoids in aphids. Insect Sci. 21:251–64.

Vorburger, C., L. Gehrer, and P. Rodriguez. 2010. A strain of the bacterial symbiont *Regiella insecticola* protects aphids against parasitoids. Biol. Lett. 6:109–111.

Vorburger, C., and A. Gouskov. 2011. Only helpful when required: A longevity cost of harbouring defensive symbionts. J. Evol. Biol. 24:1611–1617.

Vorburger, C., and S. J. Perlman. 2018. The role of defensive symbionts in host–parasite coevolution. Biol. Rev., https://doi.org/10.1111/brv.12417.

Vorburger, C., and N. Ramsauer. 2008. Genetic variation and covariation of aphid life-history traits across unrelated host plants. Bull. Entomol. Res. 98:543–553.

Vorburger, C., and R. Rouchet. 2016. Are aphid parasitoids locally adapted to the prevalence of defensive symbionts in their hosts? BMC Evol. Biol. 16:271.

Wang, W., W. Gu, Z. Ding, Y. Ren, J. Chen, and Y. Hou. 2005. A novel *Spiroplasma* pathogen causing systemic infection in the crayfish *Procambarus clarkii* (Crustacea: Decapod), in China. FEMS Microbiol. Lett. 249:131–137.

Warton, D. I., and F. K. C. Hui. 2011. The arcsine is asinine: the analysis of proportions in ecology. Ecology 92:3–10.

Watts, T., T. S. Haselkorn, N. A. Moran, and T. A. Markow. 2009. Variable incidence of Spiroplasma infections in natural populations of Drosophila species. PLoS One 4:e5703.

Wernegreen, J. J. 2002. Genome evolution in bacterial endosymbionts of insects. Nat. Rev. Genet. 3:850–861.

Werren, J. H., L. Baldo, and M. E. Clark. 2008. *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbiol. 6:741–751.

Williamson, D. L., R. F. Whitcomb, G. T. Joseph, G. E. Gasparich, D. L. Rose, P. Carle, J. M. Bové, K. J. Hackett, J. R. Adams, R. B. Henegar, et al. 1998. Revised group classification of the genus *Spiroplasma*. Int. J. Bacteriol. 48:1–12.

Woolfit, M., and L. Bromham. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. 20:1545–1555.

Xie, J., S. Butler, G. Sanchez, and M. Mateos. 2014. Male killing *Spiroplasma* protects *Drosophila melanogaster* against two parasitoid wasps. Heredity 112:399–408.

Xie, J., I. Vilchez, and M. Mateos. 2010. *Spiroplasma* bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma*. PLoS One 5:e12149.

Zchori-Fein, E., and S. J. Perlman. 2004. Distribution of the bacterial symbiont *Cardinium* in arthropods. Mol. Ecol. 13:2009–2016.

Zeng, F., G. Pederson, M. Ellsbury, and F. Davis. 1993. Demographic statistics for the pea aphid (Homoptera: Aphididae) on resistant and susceptible red clovers. J. Econ. Entomol. 86:1852–1856.

Associate Editor: K. King
Handling Editor: M. Servedio

Supporting Information

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

Figure S1. Effect of the extraction protocol on the estimated number of aphid gene copies per aphid
Figure S2. Statistical power of phylogeny in Figure 5 for detecting phylogenetic signal.
Figure S3. Rate of emergence of the three wasp lines.

Appendix chapter 2

Hugo Mathé-Hubert, Heidi Kaech, Corinne Hertaeg, John Jaenike, and Christoph Vorburger

Non-random associations of maternally transmitted symbionts in insects: The roles of drift versus biased co-transmission and selection

(2019; Molecular Ecology 28: 5330-5346; https://doi.org/10.1111/mec.15206)

Received: 9 July 2018 | Accepted: 29 July 2019
DOI: 10.1111/mec.15206

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Nonrandom associations of maternally transmitted symbionts in insects: The roles of drift versus biased cotransmission and selection

Hugo Mathé-Hubert[1] | Heidi Kaech[1,2] | Corinne Hertaeg[1,3] | John Jaenike[4] | Christoph Vorburger[1,2]

[1]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

[2]Institute of Integrative Biology, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland

[3]Institute of Agricultural Sciences, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland

[4]Department of Biology, University of Rochester, Rochester, NY, USA

Correspondence
Hugo Mathé-Hubert, Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
Email: hugomh@gmx.fr

Present address
Hugo Mathé-Hubert, CNRS UMR 5525, Laboratoire Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques, Applications, Grenoble (TIMC-IMAG), Université Grenoble Alpes, Grenoble, France

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: CRSII3_154396; US National Science Foundation, Grant/Award Number: 1144581

Abstract

Virtually all higher organisms form holobionts with associated microbiota. To understand the biology of holobionts we need to know how species assemble and interact. Controlled experiments are suited to study interactions between particular symbionts, but they only accommodate a tiny portion of the diversity within each species. Alternatively, interactions can be inferred by testing if associations among symbionts in the field are more or less frequent than expected under random assortment. However, random assortment may not be a valid null hypothesis for maternally transmitted symbionts since drift alone can result in associations. Here, we analyse a European field survey of endosymbionts in pea aphids (*Acyrthosiphon pisum*), confirming that symbiont associations are pervasive. To interpret them, we develop a model simulating the effect of drift on symbiont associations. We show that drift induces apparently nonrandom assortment, even though horizontal transmissions and maternal transmission failures tend to randomise symbiont associations. We also use this model in the approximate Bayesian computation framework to revisit the association between *Spiroplasma* and *Wolbachia* in *Drosophila neotestacea*. New field data reported here reveal that this association has disappeared in the investigated location, yet a significant interaction between *Spiroplasma* and *Wolbachia* can still be inferred. Our study confirms that negative and positive associations are pervasive and often induced by symbiont-symbiont interactions. Nevertheless, some associations are also likely to be driven by drift. This possibility needs to be considered when performing such analyses, and our model is helpful for this purpose.

KEYWORDS
coalescence, drift, symbiont-symbiont interactions, symbiotic community

1 | INTRODUCTION

Some of the interactions between organisms are so tight and durable that a new level of organisation has been defined to describe them: the holobiont (Margulis & Fester, 1991; Queller & Strassmann, 2016). These interactions are rarely bipartite and instead generally involve a host with a microbial community of varying degrees of complexity. From the host's perspective, these associations often lead to the acquisition of novel traits, allowing the host to expand its ecological niche (e.g., Brucker & Bordenstein, 2012; Henry et al., 2013; Oliver, Degnan, Burke, & Moran, 2010). Understanding the evolutionary ecology of these interactions requires identifying how

 © 2019 John Wiley & Sons Ltd

species assemble to form holobionts, both at the ontogenetic and evolutionary levels.

Large-scale screens for well-known species like *Wolbachia*, *Cardinium* or *Spiroplasma* suggest that the majority of arthropod species are infected with heritable endosymbionts (Duron et al., 2008; Hilgenboecker, Hammerstein, Schlattmann, Telschow, & Werren, 2008; Regassa, 2014; Zchori-Fein & Perlman, 2004). However, there is considerable variability in the effects these symbionts have on their hosts and in their prevalence among species. *Wolbachia* is probably the most widespread of these endosymbionts. It has been estimated to occur in 66% of arthropod species, and it typically has either low (<10%) or very high (>90%) prevalence within species (Hilgenboecker et al., 2008). *Wolbachia* is mainly known as a reproductive parasite (Werren, Baldo, & Clark, 2008), but it may also protect its host against parasites (e.g., Faria et al., 2016; Hedges, Brownlie, O'Neill, & Johnson, 2008; Teixeira, Ferreira, & Ashburner, 2008) and is sometimes necessary for successful offspring production (Dedeine et al., 2001; Kremer et al., 2009). Other widespread endosymbionts of arthropods are bacteria of the genus *Spiroplasma*, infecting 4%–7% of species, often with a low prevalence (Duron et al., 2008; Regassa, 2014), although prevalence can be high in some cases, such as in *Myrmica* ants (Ballinger, Moore, & Perlman, 2018) and in *Harmonia axyridis* (Goryacheva, Blekhman, Andrianov, Romanov, & Zakharov, 2018). Known effects of *Spiroplasma* also include reproductive parasitism (e.g., Anbutsu, Lemaitre, Harumoto, & Fukatsu, 2016; Sanada-Morimura, Matsumura, & Noda, 2013; Tabata et al., 2011) as well as defence against at least three different kinds of parasites (Ballinger & Perlman, 2017; Frago et al., 2017; Lukasik, Guo, Van Asch, Ferrari, & Godfray, 2013; Xie, Butler, Sanchez, & Mateos, 2014).

The pea aphid, *Acyrthosiphon pisum*, is one of the main biological models of endosymbiosis. It can be host to at least eight facultative heritable endosymbionts (Vorburger, 2018), including *Spiroplasma* (ixodetis clade; Fukatsu, Tsuchida, Nikoh, & Koga, 2001; Simon et al., 2011). Interestingly, Ferrari, West, Via, and Godfray (2012) showed that the communities of facultative symbionts differed strongly among host plant-associated biotypes of the pea aphid (Peccoud, Ollivier, Plantegenest, & Simon, 2009), although the prevalence of *Spiroplasma* is only weakly affected by biotype, which explains only 9% of the variance (Ferrari et al., 2012). A symbiont that experiences solely vertical transmission can persist in a host population as a reproductive parasite, or by providing a benefit to offset the cost it inflicts on the host. For example, *Spiroplasma* may protect pea aphids against entomopathogenic fungi (Lukasik, van Asch, Guo, Ferrari, & Godfray, 2013) or parasitoid wasps (Frago et al., 2017). However, this cost-benefit balance varies depending on the environment, which is thought to be the main reason for the observed polymorphism of facultative symbiont communities. For example, defensive symbioses depend on the presence of some parasites of the host, and some symbioses help the host to cope with warm environments (e.g., Russell & Moran, 2006). The cost-benefit balance may also depend on the associations with other symbionts. If two symbionts provide the same service, then one of them might be redundant and thus too costly to the host. This may be the reason why defensive bacterial symbionts are less frequent in aphids protected by ants (Henry, Maiden, Ferrari, & Godfray, 2015), or why the two defensive symbionts *Serratia symbiotica* and *Hamiltonella defensa* rarely co-occur in pea aphids (Oliver, Moran, & Hunter, 2006). Also, interactive effects between symbionts make the outcome of a given association difficult to predict. For instance, in *A. pisum*, *H. defensa* increases the titre of *S. symbiotica*, but *S. symbiotica* does not affect the titre of *H. defensa* (Oliver et al., 2006). In the presence of *Spiroplasma*, *H. defensa* decreases the fecundity of its host *A. pisum* while it increases the fecundity of the aphid *Sitobion avenae* (Lukasik, Guo, et al., 2013).

Interactive effects that vary from one symbiont strain to another limit the utility of controlled laboratory experiments, which usually include only a few particular strains, for making predictions about the overall interactions among symbionts in natural populations. For this reason, results from controlled experiments are often compared to analyses of field surveys (for several examples, see Zytynska & Weisser, 2016). These analyses notably aim at identifying pairs of symbionts for which the co-occurrence is more or less frequent than expected under the null hypothesis of random assortment (hereafter, positive and negative associations). Three kinds of mechanisms are generally considered when trying to explain such deviation from random assortment. First, the symbionts could interact in a way that increases or decreases the rate of maternal transmission failures (e.g., Rock et al., 2017), which should lead to negative or positive associations, respectively. Second, the symbionts could have an interactive effect on host fitness, enhancing or hindering their co-transmission to the next generation (e.g., Oliver et al., 2006). Thirdly, Jaenike (2012) and Smith et al. (2015) suggested a mechanism by which neutral or even slightly costly maternally transmitted symbionts could spread in the host population. These symbionts could by chance hitchhike alongside a successful symbiont whose fitness benefits outweigh the costs of the hitchhiker. Rapid spread has been reported for *Rickettsia* and *Spiroplasma* in the whitefly *Bemisia tabaci* and in *Drosophila neotestacea*, respectively (Cockburn et al., 2013; Himler et al., 2011; Jaenike, Unckless, & Cockburn, 2010). If the spreading matriline was initially associated with another symbiont as well, then faithful maternal transmission would drag it along even if it were advantageous for the host to lose the hitchhiking symbiont. This symbiont hitchhiking is analogous to genetic hitchhiking (or draft), where a neutral or slightly deleterious mutation spreads in the population because of its linkage disequilibrium with a beneficial mutation (Felsenstein, 1974). Symbiont hitchhiking might be responsible for the evolutionary maintenance of the dominant strain of the symbiont called X-type in North America. This strain is costly to its host, has not been found to provide any counterbalancing benefit, but is positively associated with the defensive symbiont *H. defensa* (Doremus & Oliver, 2017).

However, most symbionts are not strictly maternally transmitted. For example, *Rickettsia* can be transmitted via plants in whiteflies (Caspi-Fluger et al., 2012), *Spiroplasma* can be transmitted via parasitic mites in flies (Jaenike, Polak, Fiskin, Helou, & Minhas, 2007) and *Hamiltonella* can be transmitted via parasitoids in aphids (Gehrer & Vorburger, 2012). Both *H. defensa* and *Regiella insecticola* show

occasional paternal transmission (Moran & Dunbar, 2006). Jaenike (2012) argued that because of these nonmaternal transmission routes and because most symbionts show some degree of maternal transmission failure, associations due to symbiont hitchhiking should disappear rapidly. Thus, in most cases, the presence of positive (or negative) associations between symbionts should suggest an interaction that favours (or hinders) their co-occurrence. Jaenike, Stahlhut, Boelio, and Unckless (2010) showed that *Spiroplasma* and *Wolbachia* in *D. neotestacea* are positively associated despite imperfect maternal transmission. By combining these observations with a mathematical model, they suggested that these two symbionts are likely to be interacting positively with each other. As we will show in this paper, positive and negative associations are also expected to appear and persist by drift, implying that without information about the effective female population size, one needs to be cautious in assigning biological meaning to such associations.

In the first part of this study, we used a field survey of *A. pisum* symbiotic infections to identify positive and negative associations among symbionts. This analysis confirmed several previous findings that associations of symbionts often deviate from random assortment (Figure 1a). In the second part of this study, in order to understand the evolutionary meaning of these associations, we developed a model simulating the evolution of the frequency of symbiont communities in the presence of maternal transmission failures, horizontal transmissions, selection and drift. The model shows that associations of symbionts are expected to be produced by drift provided that the rates of maternal transmission failure, of horizontal transmission and the effective female population size are not too high (Figure 1b). In the third part of this study, we used the same model in the approximate Bayesian computation (ABC) framework to reanalyse the observed positive association between *Spiroplasma* and *Wolbachia* in *D. neotestacea* (Jaenike, Stahlhut, et al., 2010), combining old data (2001–2009) with new data (2010–2016). This analysis suggests that the observed dynamics of infection involve a positive interactive effect of the two symbionts on host fitness (Figure 1d).

2 | MATERIALS AND METHODS

2.1 | Natural symbiont co-occurrence

2.1.1 | Field sampling and symbiont screening

We sampled 498 aphids in France, Switzerland, Germany and Denmark during autumn 2014 and spring and summer 2015. We selected colonies that were at least 2 m apart from each other to lower the proportion of clones sampled more than once. For each sample, we recorded the host plant and the GPS coordinates. We characterised the presence of seven facultative endosymbionts by diagnostic PCR using symbiont-specific primers to amplify a part of the 16S rRNA gene (Table S1). We excluded *Wolbachia* from this analysis because of its low frequency. DNA was extracted from individual aphids using the "salting out" protocol (Sunnucks & Hales, 1996) and the PCR cycling conditions are described by Henry et al. (2013). We also ran a diagnostic PCR for the obligate endosymbiont *Buchnera aphidicola*, which is present in all aphids and thus served as an internal positive control for the quality of the DNA preparation. The nine samples that tested negative for *B. aphidicola* were excluded from the final data set. Because we had a particular interest in *Spiroplasma* infecting pea aphids (Mathé-Hubert, Kaech, Ganesanandamoorthy, & Vorburger, 2019), we also analysed the distribution of intraspecific diversity in this symbiont. This phylogenetic analysis is further described in the Supplementary material S1 and uses the strains of *Spiroplasma* described in Table S2. This analysis identified three main clades of *Spiroplasma* from pea aphids that are later referred to as clades 1, 2 and 3.

A natural population of *D. neotestacea* was sampled monthly from May through September from 2010 through 2016. During this time of year, the generation time of *D. neotestacea* is probably on the order of one month or less. Adult flies were collected by sweep netting over mushroom (*Agaricus bisporus*) baits that had been placed in a forested area in the city of Rochester, New York. Flies were screened for *Wolbachia* and *Spiroplasma* infection using the PCR methods described in Jaenike, Stahlhut, et al. (2010).

2.2 | Statistical analysis

All analyses were performed using the R software (version 3.4.4; R Core Team, 2018). Generally, associations of symbionts that are more or less abundant than expected under random assortment would be analysed using statistical tests that assume independence of observations. Our data do not fulfil this assumption as aphid samples were obtained from many different locations and dates. We thus accounted for potential spatiotemporal autocorrelation by predicting the presence or absence of symbiont species with a regression random forest model (RF). This approach is of similar efficiency as usual spatial models (Fouedjio & Klump, 2019; Hengl, Nussbaum, Wright, Heuvelink, & Gräler, 2018). In each RF explaining the presence or absence of one symbiont species in pea aphid individuals, the following explanatory variables were used: latitude, longitude, season (number of days since the start of the year), host plant on which the aphid has been sampled, aphid colour (pink or green), presence or absence of the six other symbionts (one variable per symbiont) and the total number of other symbiont species infecting the aphid. The significance of these explanatory variables was estimated using FDR adjusted *p*-values (hereafter, FDR *p*-values). The details of this analysis are described in the Supplementary material S2.

To avoid lumping together aphids of different biotypes and thus simplify the interpretation, we re-fitted these random forest models separately to aphids sampled on *Medicago sativa* and on *Trifolium* spp., which represent 30% and 33% of all field samples, respectively. We refer hereafter to these three types of models as RF_{WD} (whole data set), RF_M (*Medicago*) and RF_T (*Trifolium*). For RF_M and RF_T the host plant was removed from the set of explanatory variables. This analysis was also run to investigate the intraspecific distribution of *Spiroplasma*, by predicting, for each *Spiroplasma* infected aphid, the phylogenetic clade of *Spiroplasma* (clades 1, 2 or 3).

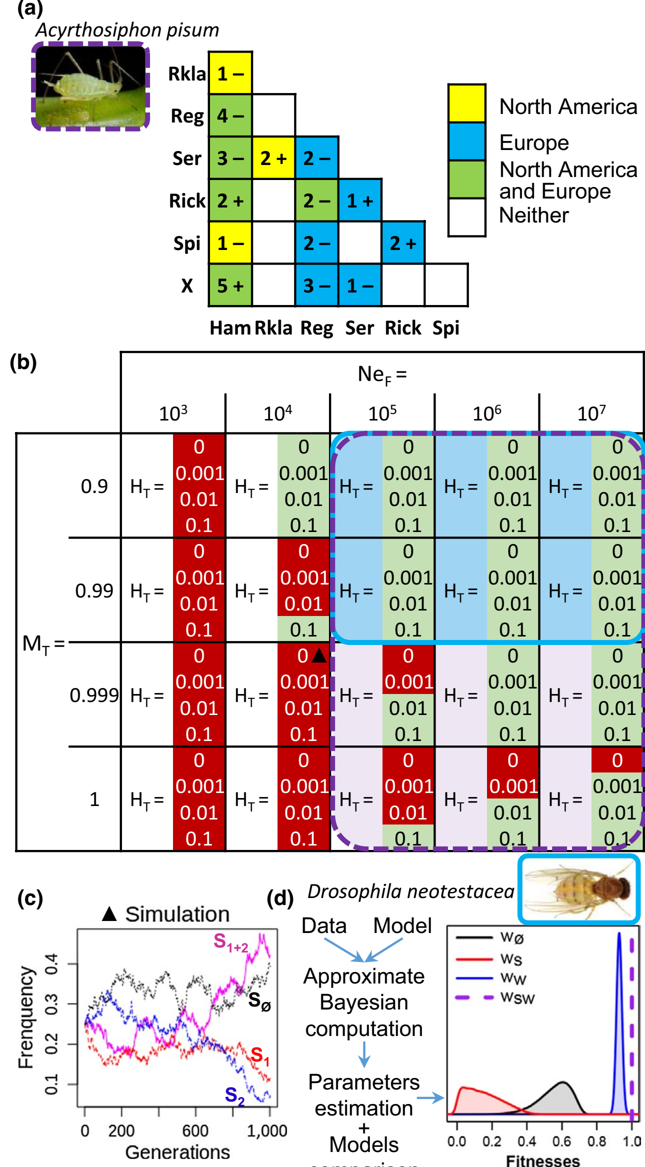


FIGURE 1 Summary of the findings. (a) For 21 pairwise combinations of pea aphid symbionts, the number of studies (including this one) that found positive (+) or negative (–) associations, and geographical locations where the associations were found. Ham, *Hamiltonella*; Rkla, *Rickettsiella*; Reg, *Regiella*; Ser, *Serratia*; Rick, *Rickettsia*; Spiro, *Spiroplasma*; X, X-type. (b) Results of the simulation analysis: combination of parameters where drift induces nonrandom assortment. Red and green values refer to combinations of parameters for which the type 1 error rate is higher and lower than 5%, respectively (based on simulated field samples of *N* = 500). Ne_F, effective female population size; M_T, maternal transmission rate; H_T, horizontal transmission rate. The blue square highlights the range of parameters values that is likely to include the symbionts of *D. neotestacea* while the dotted purple one is likely to include those of the pea aphid (Chen & Purcell, 1997; Moran & Dunbar, 2006; Peccoud et al., 2014; Rock et al., 2017). (c) Example of simulation where drift created a strong positive association. The parameters used for this simulation are pointed (▲) on the panel b: M_T, 0.999; H_T, 0; Ne_F, 10^4. (d) Analysis of the evolution of the *Spiroplasma*-*Wolbachia* association in *D. neotestacea* (Figure 6) in the approximate Bayesians likelihood framework. The density plot shows the approximate posteriors of the fitnesses of each type of fly infection relative to the fitness of flies infected by both symbionts. $w_ø$, aposymbiotic; w_s, *Spiroplasma* only; w_w, *Wolbachia* only; w_{sw}, coinfected flies. Sources for images: PLoS Biology Issue Image (2010); Werner & Jaenike (2017)

These analyses revealed that some symbionts are less frequent in aphids already containing other symbiont species, while others were not affected. To further investigate this, we characterised the link between the frequency of each symbiont species and the average number of additional symbiont species with which it co-occurs. We also investigated the effect of drift on this link. This analysis is further described in the Supplementary material S3.

2.3 | A model of evolution of symbiont co-occurrences

We developed a model of evolution of maternally transmitted symbiont co-occurrence for two purposes. Firstly, we wanted to assess the effect of drift on deviations from random assortment in the presence of various rates of maternal transmission failure and horizontal transmissions. Secondly, we used this model to analyse the well-documented case of a symbiont association between *Wolbachia* and *Spiroplasma* in *D. neotestacea*, for which estimates of the relevant parameters are available.

In short, considering only the two symbionts case, this model simulates populations of female hosts reproducing with nonoverlapping generations and being infected by zero, one or two different symbionts (species or strains). Symbionts are maternally and horizontally transmitted at varying rates. The strength of this model is that it simulates different events (reproduction and horizontal and maternal transmissions) by performing random samplings in the relevant probability distributions to update the headcount of the different types of infections, which avoids simulating every individual. The different steps for which we generate these randomly sampled values are represented by questions a–f in Figure 2. This allows the model to be fast without assuming an infinite population size. This rapidity is needed to simulate a large number of generations and replicates (simulation study), and to simulate a large number of parameter combination (ABC study). Because this model studies maternally transmitted symbionts, it only simulates females. Fitness in this model thus scales with the capacity of females to produce daughters.

With only two types of symbionts (S_1 and S_2), the population is described by the number of females being aposymbiotic ($S_ø$), having only one of the symbionts (S_1 and S_2), or having both ($S_{1,2}$). At each generation, we simulated horizontal transmissions, reproduction events and maternal transmission failures (Figure 2). The total number of horizontal transmissions is randomly chosen from a Poisson distribution whose mean depends on the horizontal transmission rate (H_T) and the frequency of the transmitted

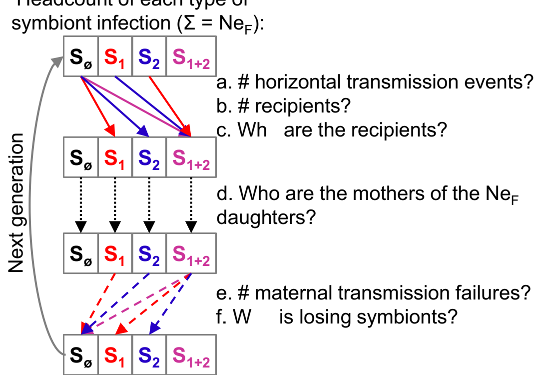


FIGURE 2 Model of evolution of symbiont associations. The population is represented as four cells corresponding to the four types of infection. Each host generation is simulated in three steps. Firstly, horizontal transmissions change some individuals from one category to the other by gaining one (or two) symbionts (solid arrows coloured as a function of the gained symbionts). This is simulated answering the questions a, b and c by randomly sampling the appropriate probability distribution. Then the reproduction is simulated by randomly choosing mothers according to the fitnesses induced by each type of infection (stippled arrows). Finally, maternal transmission failures change some individuals from one category to the other by losing one (or two) symbionts (dotted arrows coloured as a function of the lost symbionts)

symbiont (Figure 2: question a). The number of recipients can be lower than the number of horizontal transmission events when individuals receive the same symbiont more than once. The number of recipients is thus randomly chosen in a binomial distribution in which the mean depends on the number of horizontal transmissions previously drawn randomly and the total number of individuals in the population (Figure 2: question b). Finally, the repartition of the recipients among the four host classes ($S_ø$, S_1, S_2, and $S_{1,2}$) is chosen randomly from a hypergeometric distribution to simulate samplings without replacement (Rice, 2006; Figure 2: question c). Reproduction is simulated by sampling mothers from a multinomial distribution described by the headcount of females with the four kinds of symbiont communities and scaled by their relative fitness, which is determined by the fitness effects of the symbionts, which can be assumed to be multiplicative or interactive in the case of double infections. This simulates samplings with replacement (Rice, 2006; Figure 2: question d).

Maternal transmissions are simulated using the same general logic as horizontal transmissions (Figure 2: questions e and f). For more details, see the Supplementary material S4. The model is available as an R function in Appendix S1.

2.4 | Can deviations from random assortment appear by drift?

The first aim of this model was to investigate the effect of drift on symbiont associations. Hence we did not simulate any interactive effect of the symbionts on host fitness or maternal transmission, but we had to assume some noninteractive effects of the symbionts on host fitness to stabilise the polymorphism of infection, which would otherwise have disappeared rapidly under many parameter combinations (e.g., frequent maternal transmission failures or horizontal transmissions).

Specifically, we simulated 3,000 replicates of all combinations of the following sets of parameters: Effective female population sizes (Ne_F: 10^3, 10^4, 10^5, 10^6, and 10^7), successful maternal transmission rates (M_T: 1, 0.999, 0.99, 0.90), horizontal transmission rates (i.e., Average number of horizontal transmission events caused by each infected host; H_T: 0, 0.001, 0.01, 0.1). The parameter values M_T = 1 and H_T = 0 are unrealistic but were investigated to ensure that even biological systems with extremely low or high values such as associations between strains of obligatory symbionts are in the explored parameter space.

In the absence of selection, for most combinations of M_T and H_T, the symbionts get either rapidly fixed or lost. This absence of polymorphism prevents assessing deviations from random assortment. To slow down the loss of polymorphism, we set the selection on the presence of each symbiont such that it counteracts the effect of maternal transmission failures and of horizontal transmissions. The fitness of aposymbiotic hosts was set to one. Then the fitness of those infected by only one symbiont species was set to the value that, in an infinite population, would keep the frequency of the symbiont constant. For individuals infected by both symbionts, the fitness is the product of the fitnesses induced by each of its symbionts. This multiplicative fitness is similar to the model used by Jaenike, Stahlhut, et al. (2010), and corresponds to an absence of interaction between the symbionts. For more detail, see Supplementary material S5.

Populations were initiated by randomly picking the frequency of each symbiont in a uniform distribution to then set the headcount of the four kinds of symbiont communities ($S_ø$, S_1, S_2, and $S_{1,2}$) according to these frequencies and to the assumption of random assortment. The evolution of these populations of randomly assorted symbionts was then simulated for 10^5 generations or stopped if the polymorphism of infection was lost. This large number of generations was needed because the initial state of the populations, where symbionts are randomly assorted, might have actually never existed in natural populations. Therefore, the time needed for drift to induce apparent nonrandom assortment should be interpreted as an estimation of the strength of the effect of drift. This also allowed to assess the stability of deviations from random assortment once they appeared, which can take a long time in large populations.

At each generation, 500 individuals were randomly sampled from the population and used to test the significance of the deviation from the assumption of random assortment using a Chi-square test and to assess the sign of the deviation. The *p*-values were computed at every generation and recorded at generations 0, 10, 10^2, 10^3, 10^4 and 10^5. The *p*-values computed at every generation were used to assess if, as it is often assumed, associations lasting for multiple generations are unlikely to be caused by drift. We estimated the number of generations needed for a previously

significantly positive association to become significantly negative and vice-versa. This was computed for each replicate as the number of generations between the first significant deviation from random assortment and the end of the simulation divided by the number of such inversions.

2.5 | Analysing a real data set while accounting for drift

Jaenike, Stahlhut, et al. (2010) argued that *Wolbachia* and *Spiroplasma* in *D. neotestacea* are probably interacting in a way that enhances the fitness of coinfected hosts. Indeed, these two symbionts are positively associated in natural populations, despite having a maternal transmission rate of approximately 0.96, which should rapidly randomise them. We used our model in the ABC framework (Approximate Bayesian Computation) to assess how robust this conclusion is to drift. We combined the data gathered from 2001 to 2009 and analysed by Jaenike, Stahlhut, et al. (2010) with additional data gathered from 2010 to 2016 (Table S3). We fitted to these data an interaction model and a no interaction model, the latter being similar to the model of Jaenike, Stahlhut, et al. (2010). We tested for the interaction on the host fitness twice. Firstly, we looked at the interaction model and tested whether the distribution of the approximate posteriors of the interactions included zero. Secondly, we compared the quality of fit of the two models.

We used the function "ABC_rejection" of the R package "EasyABC" (Jabot, Faure, Dumoulin, & Albert, 2015) to estimate the relative fitness induced by the different kinds of infections. This approach compares observed data to the data simulated with varying values for the parameters to be estimated. To compare observed and simulated data, we assigned each field sample to one generation assuming that there were five generations per year (as Jaenike, Stahlhut, et al., 2010; details in Table S3). According to this assumption, the data set spans 77 *Drosophila* generations.

In these simulations, the parameters that are not estimated need to be fixed. These parameters are Ne_F, M_T and H_T. For M_T, we used the estimates of Jaenike, Stahlhut, et al. (2010) that range from 0.945 to 0.981. To be conservative in inferring potential interactions, we used values for Ne_F and H_T that should overestimate the effect of drift. We assumed no horizontal transmissions ($H_T = 0$) because they decrease the effect of drift (result of the simulation study). This assumption is reasonable given that a high association between the type of symbiotic infection and the mitochondrial haplotype has been observed (Jaenike, Stahlhut, et al., 2010). The effective population size (Ne) was approximately estimated using the formula $Ne = \pi/4\mu$ where μ is the mutation rate of *Drosophila melanogaster* ($\mu = 2.8 \times 10^{-9}$; 95% CI = [10^{-9}; 6.1×10^{-9}]; Keightley, Ness, Halligan, & Haddrill, 2014) and π the nucleotide diversity ($\pi = 0.0237$; 95% CI = [0.0135; 0.0337] estimated by bootstrapping autosomal loci; Pieper & Dyer, 2016). This gives an estimate of $Ne = 2 \times 10^6$, but to be conservative, we used an underestimation of the Ne using the lower CI of π and the upper CI of μ, which gives $Ne_{F_min} = 2.8 \times 10^5$, assuming a sex-ratio of 0.5.

The other parameters were estimated by randomly sampling their values from uniform priors. These parameters are the initial frequencies of *Spiroplasma* and *Wolbachia*, their initial association (measured with the phi coefficient; Everitt & Skrondal, 2010) and the fitnesses induced by the different combinations of symbionts ($w_ø$, w_s, w_w and w_{sw}). For the initial association and frequencies, the uniform prior ranged from −1 to 1 and 0 to 1 respectively. For the fitnesses, we used the same approach as Jaenike, Stahlhut, et al. (2010) which modelled the cost of not having a symbiont and took the fitness of coinfected individuals as reference by setting $w_{sw} = 1$. For the no interaction model, we estimated the fitness effect of the two other types of infected individuals (w_s and w_w) using the uniform prior ranging from 0 to 2 and we constrained $w_ø$ to be equal to $w_w \times w_s$, which assumes a multiplicative fitness effect as in Jaenike, Stahlhut, et al. (2010). For the interaction model, we also estimated $w_ø$ using the same priors as for w_s and w_w. This allowed the absence of *Spiroplasma* and *Wolbachia* to have an interactive effect on the host fitness.

For both models, the priors were randomly sampled 10^8 times. For each simulation, the randomly drawn initial symbiont frequencies and coefficient of association were used to initiate the population whose evolution was simulated by the model for 77 generations according to randomly drawn fitness effects of symbionts and the fixed parameters (Ne_F, H_T, M_T). The two sets of 10^8 simulated data sets were summarised and compared to the summary of the observed data set. These summaries contain the mean frequencies of the four types of infections at start, midpoint and end (details in Table S3). Simulations were "accepted" and used to estimate parameters when the Euclidean distance between their summary and the summary of the observed data was below the tolerance threshold of 0.153. This tolerance was chosen to accept at least 1,000 simulations per model, which is 0.001% of the simulations.

We estimated the cost of not having *Wolbachia*, *Spiroplasma* or their synergetic effect by applying a similar formalism as Jaenike, Stahlhut, et al. (2010) to the distribution of the approximate posteriors of the fitnesses. Jaenike, Stahlhut, et al. (2010) modelled the cost of not having a symbiont by setting $w_{sw} = 1$; $w_s = 1 - s_w$; $w_w = 1 - s_s$; $w_ø = (1 - s_w) \times (1 - s_s)$, where s_s and s_w are the cost of not having *Spiroplasma* or *Wolbachia*, respectively. This corresponds to the situation modelled by the no interaction model, while for the interaction model, we extended this formalism by setting $w_ø = (1 - s_w) \times (1 - s_s) \times (1 - s_{sw})$, where s_{sw} is the cost of not having the synergetic effect between *Spiroplasma* and *Wolbachia*.

These fitnesses and costs, as well as the initial state of the population, were estimated using the mode of posterior distributions of the interaction and no interaction models, and their 95% confidence interval using the 0.025 and 0.975 percentiles. We performed a pairwise comparison of the estimated fitness effect of the four types of infections. For each pair of infection type we tested the significance of the differences by subtracting their approximate posterior distributions and assessing the extent to which the resulting distribution overlaps with zero. Specifically, we

tested the null hypothesis $w_1 = w_2$ using the two-sided Bayesian p-value: $\begin{cases} 2 \times \text{Freq.}(w_1 < w_2) & \text{if } \overline{w_1} > \overline{w_2} \\ 2 \times \text{Freq.}(w_1 > w_2) & \text{if } \overline{w_1} < \overline{w_2} \end{cases}$ as where the frequencies (Freq.) are estimated over the posteriors (i.e., the "accepted" simulations). We further tested the interaction by comparing the quality of fit of the two models through the delta of the Bayesian predictive information criterion (BPIC; Ando, 2007; Turner, Sederberg, & McClelland, 2014).

3 | RESULTS

3.1 | Natural co-occurrence of pea aphid symbionts

The random forests (RFs) model analysing the associations of symbionts in field-sampled aphids revealed three positive associations and six negative associations. Of these associations, all were detected in the whole data set (RF_{WD}, Figure 3a) six were detected in aphids from *Trifolium* spp. (RF_T, Figure 3b), and only two were detected in aphids from *M. sativa* (RF_M, Figure 3c). The sample size and the average number of symbionts per aphid were similar in the *M. sativa* (RF_M) and in the *Trifolium* spp. (RF_T) group (*M. sativa*: 148 aphids with 0.97 symbionts per aphid on average; *Trifolium*: 161, 0.77). Therefore, it is unlikely that the lower number of significant associations in *M. sativa* (RF_M) is caused by lower statistical power. Of the 11 significant associations already identified by other studies on pea aphids, six were also found in this study, and all associations reported by several studies (including ours) were always of the same sign (Table 1; Figure 1a). Particularly noteworthy are the consistently negative associations between the common symbionts *H. defensa* and *R. insecticola*, and the consistently positive associations between *H. defensa* and X-type.

To account for the nonindependence between samples, these models included the variables longitude, latitude, season and host plant (only RF_{WD}). However, these variables are highly correlated, and although we used conditional inference trees and conditional importance, the results should be interpreted with caution. The effects of these four variables on the frequency of each symbiont are described in Figure S1.

Some symbiont prevalences covaried negatively with the total number of coinfecting symbiont species, whatever their identity (*H. defensa*: FDR p-values = .002 and .02 in RF_{WD} and RF_M, respectively; *R. insecticola*: FDR p-value < .001 in the three models RF_{WD}, RF_T and RF_M; *S. symbiotica*: FDR p-values < .001 and in both RF_{WD} and RF_M; Figure 4). For pea aphids from *Trifolium* spp., the relationship between symbiont prevalence and the mean number of coinfecting symbionts was tight (R^2 = 0.98). The slope was more negative than −1 which is the slope expected under random assortment (slope = −2.16; p-value < .001, Figure 4b). This observation was mostly driven by *R. insecticola*. However, repeating the analysis without aphids infected by *R. insecticola* did not change the result much (R^2 = 0.88; slope = −2.26; p-value = .003). For pea aphids from *M. sativa*, there was no detectable relationship between the frequency of symbionts and the number of coinfecting symbiont species (R^2 = 0.14; slope = −1.61; p-value = .35, Figure 4a). The simulations described in Supplementary material S3 revealed that this relationship is also affected by drift, which increases variation in the slopes around the expected value of −1 and moderately decreases the proportion of variance explained.

3.2 | *Spiroplasma* intraspecific diversity

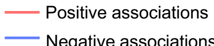
The phylogenetic tree indicates that in Europe, pea aphid infecting *Spiroplasma* are subdivided into at least three clades, although clade 3 has low bootstrap support (Figure S2). The relative frequencies of these three clades did not depend on the host plant (p-value = .98; Figure S2) but were strongly dependent on the symbiont community. Clade 2 was more frequent in aphids already infected by other endosymbionts (FDR p-value = .01) than the other two clades. The difference of clade 2 to clade 1 was marginally nonsignificant, while the difference to clade 3 was marginally significant (p-values = .06 and .03, respectively; Wilcoxon-test). The *Spiroplasma* clades were also differently associated with *H. defensa*, X-type and *Rickettsia* (FDR p-values = .02, .003 and .003, respectively). Specifically, clade 3 co-occurs less frequently with *H. defensa* than clades 1 and 2 (p-values = .02 and .01; Fisher-exact test) and more frequently with X-type than clades 1 and 2 (p-values = .003 and .006; Fisher-exact test; Figure 3 and Figure S2). Also, clade 2 is more frequently associated with *Rickettsia* than clades 1 and 3 (p-values < .001 in both cases; Fisher-exact tests; Figure 3 and Figure S2).

3.3 | Simulations of the symbiont co-occurrences evolving by drift

Symbiont associations that are more or less frequent than expected under random assortment are generally thought to be the signature of an interaction between the symbionts that promotes or prevents their co-occurrence. Our simulations showed that when M_T = 1 and H_T = 0, drift always leads to strong deviations from random assortment, although associations take longer to establish in large populations where drift is weak (Figure 5). As expected, less-than-perfect maternal transmission or horizontal transmission tend to randomize symbiont associations (Figure 5 and Figure S3). However, our model shows that this effect can be offset by drift, in particular under effective population sizes lower than 10^6 (Figure 5). For effective female population sizes (Ne_F) of 10^3, 10^4 and 10^5 or more, it takes a median number of 54, 117, and 211 generations to inverse the sign of a significant deviation from random assortment (Figure S4). Symbiont associations due to drift alone can thus be quite persistent in time.

3.4 | *Spiroplasma-Wolbachia* association in *D. neotestacea*: Drift or selection?

The positive association between *Spiroplasma* and *Wolbachia* in *D. neotestacea* reported in Jaenike, Stahlhut, et al. (2010) has declined slowly from 2001 to 2016 and now seems absent. The frequency



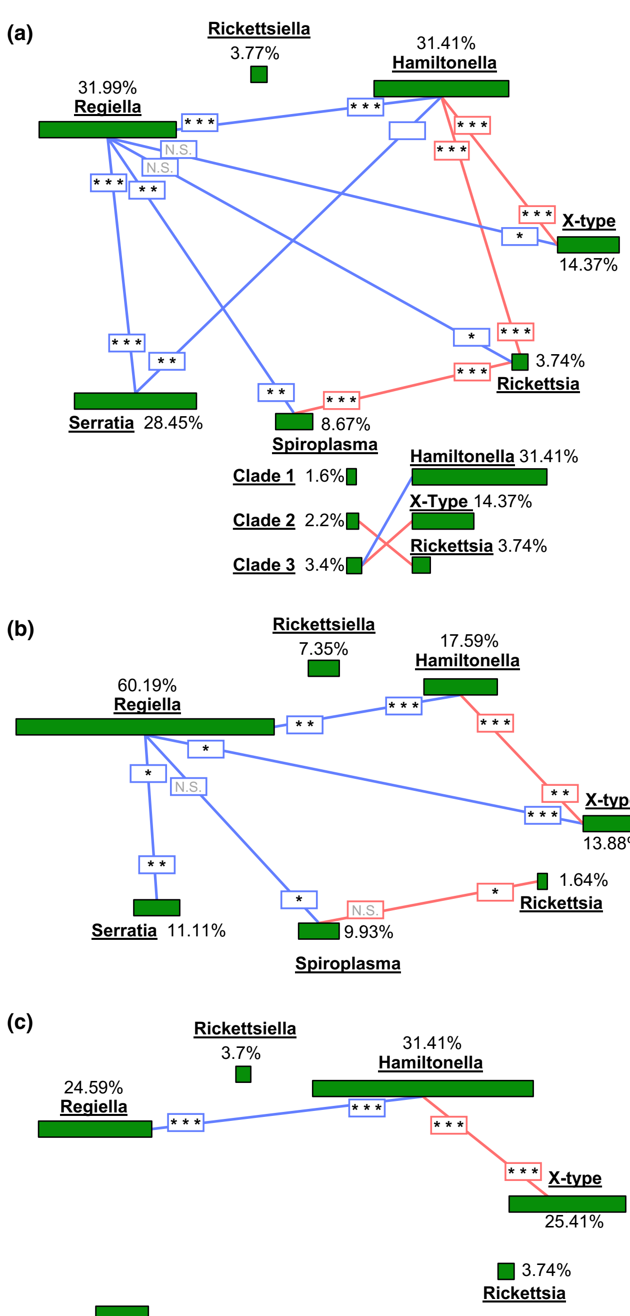


FIGURE 3 Patterns of symbiont co-occurrence. The seven symbiont species are represented by green boxes whose size is proportional to the overall prevalence of the symbiont in the whole data set (a; $N = 498$), in aphids from *Trifolium* spp. (b; $N = 161$) and in aphids from *Medicago sativa* (c; $N = 148$). Red and blue lines connect symbionts that co-occur more or less often than expected under random assortment, respectively. Stars indicate the FDR-adjusted level of significance of these associations and are placed close to the symbiont that was the dependent variable in the random forest models

of coinfected flies has shifted from approximately 0.75 to 0.4 while the frequency of flies infected by *Wolbachia* only has shifted from approximately 0.2 to 0.5 (Figure 6). The disappearance of the positive association calls into question the previous conclusion of an interactive fitness effect of these two symbionts. However, when we compare a no interaction and an interaction model fitted to these data, we do indeed find support for a positive interaction on host fitness. The relationship between the values of the parameters and the distance between the summaries of the simulated and observed data sets are shown in Figure S5. The estimated parameters revealed a clear interaction since the fitnesses of the four host categories had the following rank order: $w_s < w_ø < w_w < w_{ws} = 1$ (*p*-values = .006, <.001, <.001 respectively; Figure 1d; estimated values are in Table 2). This sorting resulted in a cost of not having the synergetic effect of *Spiroplasma* and *Wolbachia* that is twice higher than the cost of not having the beneficial effect of *Wolbachia* ($s_{ws} = 0.78$ vs. $s_w = 0.38$; interaction model in Table 2). This interaction is also demonstrated by a delta of Bayesian predictive information criterion of 7.75. The interaction model also revealed that infection by *Spiroplasma* would actually be costly to the host, since the cost of not having *Spiroplasma* was negative ($s_s = -3.13$, 95% CI = [−90.4; −0.36]; Table 2). Consistent with this result, strains of *D. neotestacea* only infected by *Spiroplasma* can be difficult to maintain in the laboratory (John Jaenike, personal observation).

4 | DISCUSSION

Understanding how symbionts associate and interact within a host is important but challenging. Laboratory experiments address this question by controlling all relevant parameters and observing the outcomes, but they can only accommodate a tiny portion of the natural diversity of each interacting species. In addition, such studies have often found that the outcome depends on the genotypes of the interacting partners (e.g., Hansen, Vorburger, & Moran, 2012; Lukasik, Asch, et al., 2013; Niepoth, Ellers, & Henry, 2018; Oliver, Degnan, Hunter, & Moran, 2009; Russell & Moran, 2006; Vorburger & Gouskov, 2011; Weldon, Strand, & Oliver, 2013), further complicating general predictions about these interactions in natural populations. Comparisons with field observations are therefore essential. When analysing field surveys, interactions between symbionts are tentatively inferred by comparing the observed frequency of co-occurrences to the frequency expected under the hypothesis of random assortment. Departures from random assortment have been reported frequently in pea aphids. Indeed, of the 21 possible pairwise associations among the seven facultative endosymbionts considered here, 11 have already been reported to have significantly higher or lower frequencies than expected under random assortment in earlier studies on pea aphids (Figure 1a and Table 1). Six of these associations were also found in our field sampling, and three are reported for the first time. When focusing on *Spiroplasma*, we even found significant associations at the intraspecific level. The three main *Spiroplasma* clades identified in the phylogenetic tree

TABLE 1 Patterns of symbiont co-occurrence in this study and in other studies on pea aphids

	RF_{WD}	RF_T	RF_M	Oliver et al. (2006)	Rock et al. (2017)	Ferrari et al. (2012)		Russell et al. (2013)	Henry et al. (2013)
Host plant	Many	T	M	M	M	T	V	T V M	Many
Geographic location	E.	E.	E.	E.	N.A.	E.	E.	N.A.	N.A.& E. (14 countries)
Regiella/*Serratia*	–	–							
Regiella/*Spiroplasma*	–	–							
Regiella/*Rickettsia*	–				–				
Regiella/X-type	–	–				–			
Regiella/*Hamiltonella*	–	–	–			–		–	
Serratia/*Rickettsia*						+			
Serratia/X-type							–		
Serratia/*Hamiltonella*	–			–	–				
Serratia/*Rickettsiella*					+			+	
Spiroplasma/*Rickettsia*	+	+							
Spiroplasma/*Hamiltonella*					–				
Rickettsia/*Hamiltonella*	+				+				
X-type/*Hamiltonella*	+	+	+		+				+
Hamiltonella/*Rickettsiella*					–				

Note: E, Europe; M, *Medicago sativa*; N.A., North America; T, *Trifolium* spp.; V, *Vicia*.

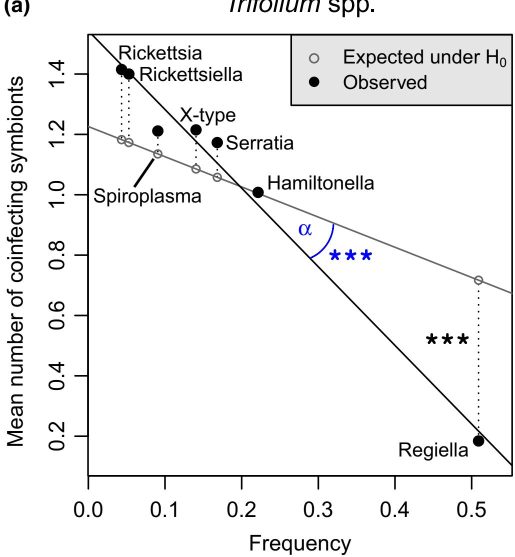
were nonrandomly associated with other symbionts, independent of the host plants the aphids were collected from. Such intraspecific variation in a symbiont-symbiont association has also been reported between X-type and *H. defensa* in the pea aphid (Doremus & Oliver, 2017). But what is the biological meaning of these pervasive associations?

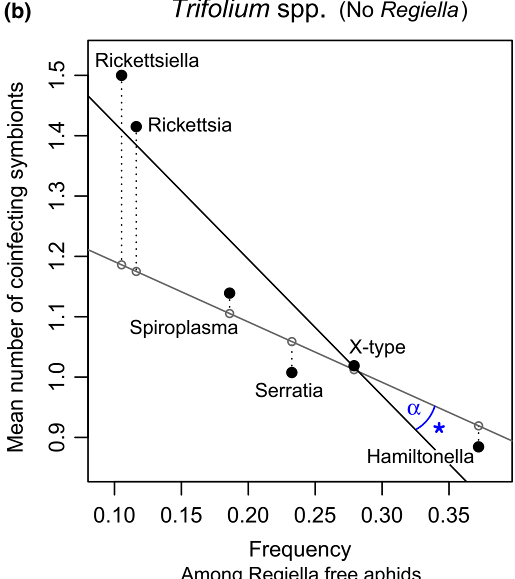
4.1 | Drift induces deviations from random assortment

Our simulation model showed that, albeit a random phenomenon, drift alone can induce associations among maternally transmitted symbionts, suggesting that random assortment is not an appropriate null model to compare symbiont coinfections against. The reason is most easily understood by considering the coalescence framework. Statistical tests used to detect departures from random assortment assume that samples are independent of each other. While this may apply to horizontally transmitted symbionts, it will not apply to maternally transmitted symbionts. Some individuals will have the same symbiont association simply because they share a female ancestor that transmitted this particular symbiont community to all of its offspring. In population genetics, this phenomenon is referred to as coalescence (Balding, Bishop, & Cannings, 2007), which should not be confounded with the "community coalescence" (Rillig et al., 2015). One of the measures of the strength of drift is the expected coalescent time, the average number of generations between two randomly sampled alleles and their most recent common ancestor. It is equal to 2Ne for diploid autosomal genes, but it is only Ne/2 for maternally transmitted cytoplasmic genomes (assuming a sex-ratio of 0.5). This is because only females transmit the cytoplasmic genome, and they have only one copy of it (Jaenike, 2012; Moore, 1995). Cytoplasmic genomes, including endosymbionts, hence undergo four times more drift than nuclear autosomal genes.

Jaenike (2012) investigated how the population genetics framework can be adapted and used to study the evolution of communities of maternally transmitted symbionts by comparing each symbiont to a gene. However, given the generally high fidelity of maternal transmission and the low rate of horizontal transmission of endosymbionts, one could also compare the whole symbiont community to one gene with many alleles. Mutations increase allelic diversity, while drift has the opposite effect. This mutation-drift equilibrium is largely analogous to the balance between maternal transmission failures, horizontal transmissions and drift that we studied with our model. The main difference is that maternal transmission failure effectively acts as a directional mutation pressure, where the number of individuals mutating from one state (infected) to the other (uninfected) is proportional to the number of individuals in the original state (infected), which is not true for horizontal transmission. The probability of undergoing horizontal transmissions increases with the frequency of the symbiont, which makes polymorphism less easily maintained in the presence of horizontal transmission.

Drift-induced deviations from random assortment can persist for a very long time. In a population of diploid autosomal genes, a neutral mutation that reaches fixation does so, on average, 4Ne generations





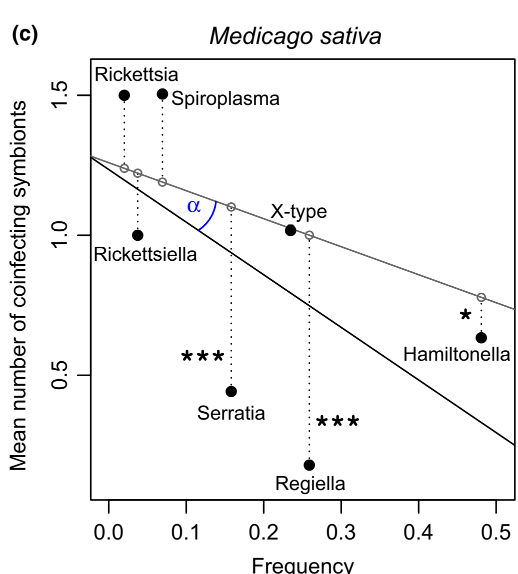


FIGURE 4 Relationship between symbiont frequency and mean number of other symbionts species. Comparison of the actual (black) and expected (grey) relationship between the frequency of endosymbiont species and the mean number of other symbiont species with which they co-occur. Each observed value is connected to its expected value by a dotted line. Stars along these lines indicate the FDR adjusted level of significance detected by random forest models. Analysis was performed on pea aphids from *Trifolium* spp. (a and b) and *Medicago sativa* (c). Panel b refers to the analysis performed on aphids from *Trifolium* spp., but excluding individual infected with *Regiella insecticola* from the analysis. For each of these three cases, we tested if the angle between the two slopes (α) differed significantly from zero [Colour figure can be viewed at wileyonlinelibrary.com]

after it appeared (Kimura & Ohta, 1969), or after Ne generations in a haploid, maternally transmitted gene. Thus, we should expect that drift-induced deviations from random assortment of symbionts should also be somewhat stable in time. In agreement with that, our simulations of two strictly maternally transmitted symbionts show that drift-induced inversions of the sign of significant deviations from random assortment occur every 50–200 generations on average, depending on the effective female population size. These numbers should not be used as a general reference, however, because significance depends on the size of the samples used to assess deviations from random assortment (500 hosts in our simulations). Departures from random assortment became less stable in the presence of horizontal transmissions and maternal transmission failures.

4.2 | *Spiroplasma-Wolbachia* association in *D. neotestacea*: Drift or selection?

Jaenike, Stahlhut, et al. (2010) studied the maintenance of the positive association between *Wolbachia* and *Spiroplasma* in *D. neotestacea*. They used a deterministic mathematical model to show that given the maternal transmission rate estimated at 0.96, the association should disappear very rapidly in the absence of positive interactions between the two symbionts. While it is true that this relatively imperfect maternal transmission will push a population towards random assortment, their model only considered the frequency of the symbionts. Thus, it implicitly assumed an infinite population size and omitted drift which, as we have shown, pushes populations towards nonrandom assortment. The additional data collected since this study revealed that, at least in Rochester NY, the association has disappeared. Specifically, the frequency of coinfected flies decreased while the frequency of flies only infected by *Wolbachia* increased.

At a first glance, the disappearance of the association seems to reinforce the view that it could have been driven by drift. However, considering that the effective female population size is probably above 2.8×10^5, and the maternal transmission rate below 0.99, our simulation study revealed that drift alone is unlikely to induce such significant deviations from random assortment as they have been observed between 2001 and 2009. As discussed by Jaenike (2012), such associations could be driven by symbiont hitchhiking, if one of

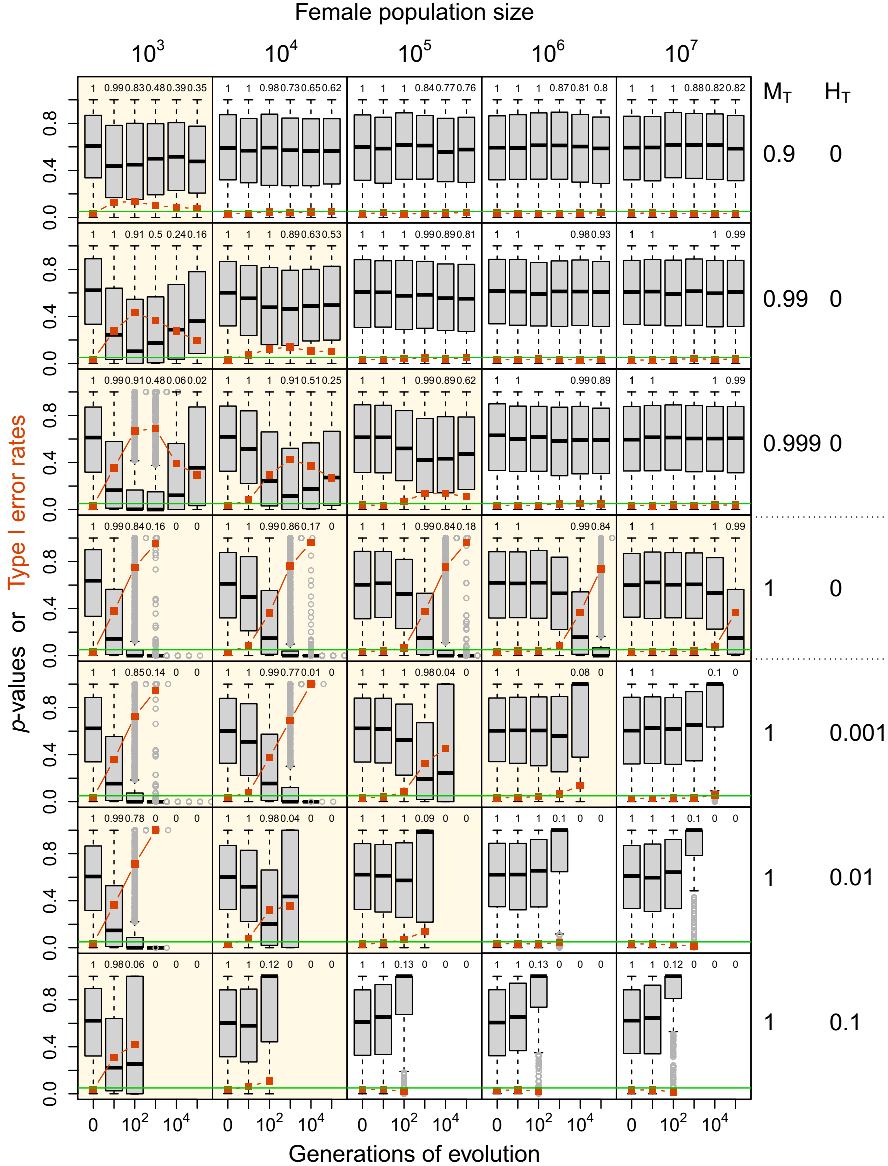


FIGURE 5 Deviations from random assortment induced by drift. The frequency of two maternally transmitted symbionts evolved for up to 10^5 generations, starting from a population in which symbionts were randomly assorted. Boxplots show the *p*-values of Chi-square tests assessing the deviation from random assortment at generations 0, 10, 10^2, 10^3, 10^4 and 10^5. Each set of boxplots corresponds to 3,000 populations evolving with the combination of the parameters indicated on the side: female population size (columns), horizontal transmission rate (H_T, rows) and maternal transmission rate (M_T, rows). The green horizontal line shows the 0.05 threshold, and the orange squares and lines indicate the type 1 error rate. Analyses of field surveys testing for deviation from random assortment usually assume that the type 1 error rate is 0.05. Combinations of parameters where this is not the case have a yellowish background. The numbers above the boxplots indicate the proportion of populations that still retained a polymorphism of infection by both symbionts [Colour figure can be viewed at wileyonlinelibrary.com]

the two symbionts is beneficial and spreads in the population from a matriline also containing another symbiont. *Spiroplasma* has actually undergone such a spread (Cockburn et al., 2013; Jaenike, Unckless, et al., 2010), probably because of the protection it provides against the parasitic nematode *Howardula aoronymphium* (Jaenike, Unckless, et al., 2010). This spread could strongly decrease the female effective population size, which was only partially accounted for in our ABC analysis since we assumed that males and females have the same effective population sizes. On the other hand, we used a conservatively low estimate of the effective female population size and the analysis still supported a strong interactive effect of *Spiroplasma* and *Wolbachia* on host fitness. Indeed, in the presence of *Wolbachia*, *Spiroplasma* infected flies had the highest estimated fitness while in the absence of *Wolbachia* they had the lowest estimated fitness. Importantly, we did not estimate parameters explaining the initial association but parameters explaining the evolution of the association. Thus this interactive fitness effect is not deduced from the presence of the association, which could have been due to symbiont hitchhiking, but from the dynamic of its disappearance, which was slower than expected given the relatively high rates of maternal transmission failures (Jaenike, Stahlhut, et al., 2010). This analysis also revealed that whatever the presence of *Spiroplasma*, *Wolbachia* always increases the fitness of its host. A more unexpected result of this analysis is that in the absence of *Wolbachia*, infection with *Spiroplasma* is inferred to be costly to the host. This estimated cost contrasts with the result of Jaenike, Unckless, et al. 2010), that *Spiroplasma* is beneficial by protecting its host from the sterilising effect of the parasitic nematode *H. aoronymphium*, while having no detectable effect on the egg count per ovary.

This surprising result of the ABC analysis results from the fact that the frequency of flies infected only by *Wolbachia* increased while the frequency of flies infected only by *Spiroplasma* remained

TABLE 2 Parameters estimated by the ABC analysis fitting the model. Three kinds of parameters were estimated; the initial population state, the fitnesses corresponding to the different types of symbiont infections, and the corresponding costs of not having a symbiont. The cells in grey correspond to parameters that were estimated by solving the equations shown in the second column of the table (see Materials and methods). The 95% confidence intervals of the parameters are given in brackets

	No interaction BPIC = 25.31	Interaction BPIC = 17.56
Initial population		
f_S	0.64 (0.61; 0.68)	0.66 (0.61; 0.72)
f_W	0.64 (0.60; 0.69)	0.66 (0.59; 0.73)
Phi	0.70 (0.62; 0.78)	0.67 (0.54; 0.78)
Fitnesses (w_{sw} = 1)		
$w_ø = (1 - s_w) \times (1 - s_s) \times (1 - s_{ws})$	0.29 (0.10; 0.41)	0.57 (0.34; 0.69)
$w_s = (1 - s_w) \times (1 - s_{ws})$	0.31 (0.11; 0.45)	0.13 (0; 0.37)
$w_w = (1 - s_s) \times (1 - s_{ws})$	0.93 (0.92; 0.93)	0.93 (0.92; 0.93)
Costs of not having symbionts		
s_s	0.07 (0.07; 0.08)	−3.13 (−90.4; −0.36)
s_w	0.68 (0.55; 0.89)	0.38 (0.25; 0.63)
s_{sw}	Set to 0	0.78 (0.32; 0.99)

constant. We assumed that the imperfect maternal transmissions estimated by Jaenike, Stahlhut, et al. (2010) are exact and representative of the considered time series. According to this assumption and in the absence of selection, maternal transmission failures would convert coinfected flies into flies only infected by *Wolbachia* or by *Spiroplasma* (at rates of 3% and 4%, respectively) and these flies would be converted into aposymbiotic flies (at a rate of 5% and 2%, respectively). With these conversion rates and an initial coinfection frequency of 60%, about 2.4% of flies should become infected by *Spiroplasma* only every generation, yet such flies remained at a constantly low frequency, revealing the cost of *Spiroplasma* in the absence of *Wolbachia*.

This cost of *Spiroplasma* contrasts with its known protective effect, which is conditional on the presence of the parasitic nematode. Possibly, *Spiroplasma* has some fitness costs that are not detected through the egg count per ovary used as a fitness proxy by Jaenike, Unckless, et al. (2010). Nevertheless, we should also consider that this result could arise from some of the necessary approximations in our analysis. For example, we considered that the rates of successful maternal transmissions estimated by Jaenike, Stahlhut, et al. (2010) were constant over time. However, the maternal transmission rate is under strong selection and it can vary with temperature, which could have influenced our inferences. This highlights that the ABC approach applied here can be useful to test hypotheses on field data, but the resulting parameter estimates must be interpreted cautiously.

Another assumption we made is the absence of horizontal transmissions. This assumption is reasonable given the high association observed between the infection status of the flies and their

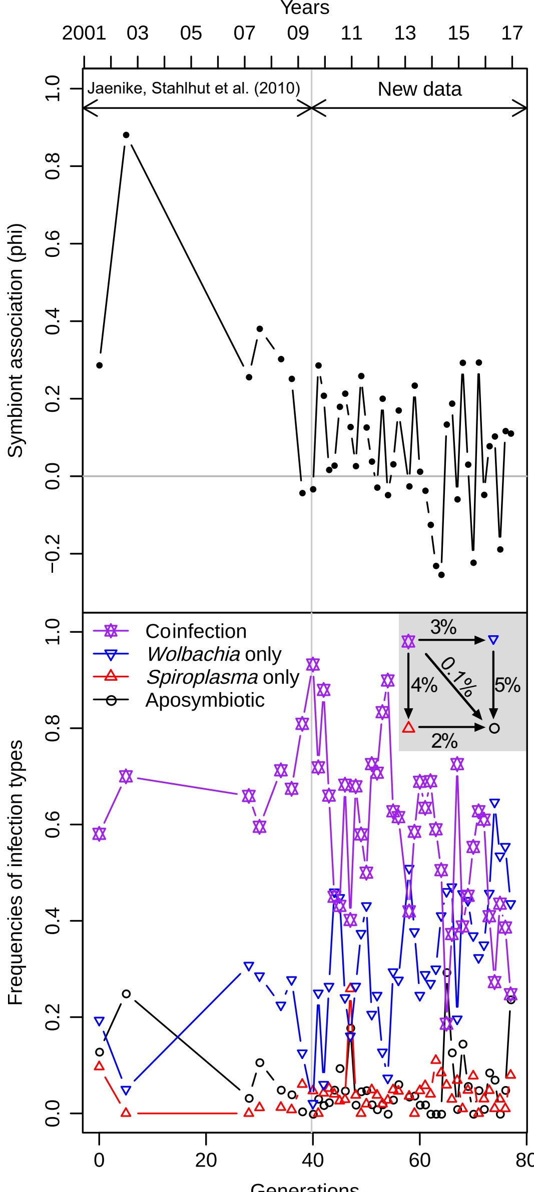


FIGURE 6 *Spiroplasma-Wolbachia* association in *D. neotestacea* in Rochester NY. The evolution of the symbiont association is shown on the upper panel while the frequencies of the four kinds of associations are shown on the lower panel. The time in years is shown at the top and the corresponding number of generations at the bottom. The diagram in the top right corner of the lower panel describes the effect of maternal transmission failures on the rates of conversion between the four types of infections. These rates were estimated by Jaenike, Stahlhut, et al. (2010) and these data were combined in the ABC framework to estimate the fitness effect of the four types of infection [Colour figure can be viewed at wileyonlinelibrary.com]

mitochondrial haplotype (Jaenike, Stahlhut, et al., 2010). This knowledge has strongly increased the statistical power of our analysis. For this reason, for any study that would plan to use such ABC approach to infer symbiont-symbiont interaction from field surveys, we would recommend to also sequence the COI gene. Then haplotypes could be included in the analysis by considering them as a symbiont

species with a known perfect maternal transmission and a null rate of horizontal transmission. With such a setting, our model could also estimate the rates of horizontal transmissions of symbionts or paternal inheritance.

Fromont, Adair, and Douglas (2019) have found in a Rochester, NY population of *D. neotestacea* that the density of *Wolbachia* did not differ significantly between *Spiroplasma*-infected and uninfected flies, whereas the density of *Spiroplasma* was positively and significantly correlated with that of *Wolbachia* among coinfected flies. Together, these findings suggest that *Wolbachia* has a positive effect on *Spiroplasma* density, and thus perhaps on maternal transmission fidelity, but that *Spiroplasma* does not have such an effect on *Wolbachia*. *Wolbachia* benefits from the presence of *Spiroplasma* because of the latter's restoration of fertility in nematode-parasitized females.

4.3 | Symbiont associations in pea aphids: Selection or drift?

After emphasizing the importance of considering drift as a source of nonrandom assortment among symbionts, we return to the interpretation of positive and negative associations among facultative endosymbionts observed in pea aphids. Are they maintained by interactions among symbionts or just a consequence of drift? Good estimates of effective female population size would obviously help. Unfortunately, this is a tricky problem in aphids and other cyclical parthenogens. Although aphids can reach enormous population sizes, they undergo a bottleneck each winter, and clonal selection during the asexual phase of the life cycle (approximately 7–14 generations in pea aphids; Barker, 2016) can be intense (e.g., Vorburger, 2006), which will also reduce the effective population size. This clonal selection acts on the three components of genetic variance (additive, epistatic and of dominance), but the optimisation it induces on the nonadditive variances is lost at each sexual generation, which maintains the presence of clonal selection from year to year (Lynch & Deng, 1994). On the other hand, aphids are good dispersers and exhibit shallow genetic population structure over large geographic scales. For example, Ferrari et al. (2012) reported F_{ST}-values ranging from 0.03 to 0.11 for pea aphid populations from the same host plants across different European countries, and Via and West (2008) reported a mean F_{ST} of 0.03 for North American populations of the pea aphid. Such high population connectivity should have a positive effect on effective population size. We do not know the effective population size of pea aphids, but DNA sequence-based estimates from other cyclical parthenogens, waterfleas of the genus *Daphnia*, are rather high (300,000–600,000; Haag, McTaggart, Didier, Little, & Charlesworth, 2009). If estimates were similarly high for pea aphids, the importance of drift in generating nonrandom assortment of symbionts would be limited (Figure 5).

Another important aspect to consider is the consistency of the sign of significant associations. While drift will generate associations of random and (slowly) fluctuating sign, selection is expected to consistently favour either positive or negative associations between particular pairs of facultative endosymbionts. For significant associations that were discovered in multiple studies, the sign of the association was always the same (Table 1). Finding particular combinations of symbionts consistently over- or underrepresented across different times and places suggests they are not caused by drift. For example, the European pea aphids population is thought to have colonised North America at least 200 years ago, which would represent 1,400–2,800 pea aphid generations, and there is strong genetic differentiation among pea aphids from the two continents today (Brisson, Nuzhdin, & Stern, 2009). Despite this separation, the four associations that have been reported in both continents are of the same sign. This strongly suggest that at least some of them are driven by an interaction between the symbionts. Indeed, if these associations were inherited from the pea aphids that invaded North America, then it has been stable for more than 1,000 generations, which is unlikely for associations driven by drift (Figure S4).

In addition to testing for deviations from random assortment, some studies have also assessed whether symbiont species tend to be differently associated with aphids that are already infected with 0, 1, 2 or more other symbiont species (e.g., Ferrari et al., 2012; Rock et al., 2017; Russell et al., 2013; Zchori-Fein, Lahav, & Freilich, 2014). In our field survey, we found that *H. defensa*, *S. symbiotica* and *R. insecticola* occurred more frequently in aphids containing no or few other symbiont species than expected under the assumption of random assortment, although this was only significant in aphids sampled from *M. sativa*.

We further investigated this by characterising the link between the frequency of symbionts and the number of coinfecting symbiont species. This link is expected to be strong because frequent symbionts are less likely to share a host with other symbiont species than rare symbionts, leading to an expected slope of −1 (Supplementary material S3). This reveals that rare symbionts are more strongly selected to cope with other symbiont species than abundant symbionts (this is also true for horizontally transmitted symbionts). We found that this slope was nonsignificant in aphids sampled on *M. sativa*, and significantly lower than the expected value (−1) in aphids sampled on *Trifolium* spp. (Figure 4). These results might be the consequence of drift, constraints, or adaptations. For example, rare symbionts might be rare because they need the presence of other symbionts to persist in the host population. Such constraint would reinforce the expected relationship. Alternatively, since rare symbionts are expected to co-occur on average with more symbiont species than abundant ones, these rare symbionts might have become better adapted to the presence of other symbiont species, thus reinforcing the expected pattern. This highlights that only abundant symbiont associations are efficiently optimised by natural selection. It is therefore worth considering that associations between symbionts that are currently maintained by a positive interaction may have evolved as a consequence of an association that had initially appeared by drift or hitchhiking.

Lastly, inference on the biology of particular symbionts or their associations can be strengthened from analyses of seasonal patterns

and their comparison with expectations from laboratory experiments. In studies of seasonal dynamics, the effect of drift is ideally ruled out using spatiotemporal replication. For example, Smith et al. (2015) reported a correlated change in the symbiont frequencies and the parasitoid-induced host mortality which, together with the laboratory evidence for symbiont-conferred resistance against parasitoids, suggested a causal relationship between them. Also, Montllor, Maxmen, and Purcell (2002) reported an increase in the frequency of *S. symbiotica* correlated with temperature, which was consistent with this symbiont helping to tolerate heat stress. Our sampling design was not suited for such inference, but the result that *H. defensa* was more abundant in summer than in spring (Figure S1) was at least consistent with selection by parasitoids as also reported by Smith et al. (2015). Field observations are also informative when they do not match the expectations from laboratory work. For example, laboratory experiments suggested that X-Type does not provide any detectable benefit to the pea aphid, but it is quite frequent and positively associated to *H. defensa*, suggesting it might have benefited from hitchhiking during the spread of *H. defensa* (Doremus & Oliver, 2017). Also, Wulff, Buckman, Wu, Heimpel, and White (2013) did not find that the symbiont *Arsenophonus* was protecting its *Aphis glycines* host against its main parasites, but it was present at high frequency. This discrepancy between observation and expectation motivated further experiments revealing that *Arsenophonus* provides a general – et to be described – benefit to the aphid (Wulff & White, 2015). Although difficult to interpret, field surveys remain crucial for our understanding of the ecology of symbioses.

In conclusion, the fate of holobionts depends on host-symbiont interactions as well as on symbiont-symbiont interactions, but identifying them is not always straightforward. The approach consisting in analysing the frequency of associations in the field is useful. However, the results it yields must be interpreted carefully, particularly in the case of maternally transmitted symbionts, as patterns expected to be produced by interactions between symbionts are also induced by drift. The model we developed can help this task. The study of this model highlights that holobionts are not only a source of additional units of selection, but also a source of additional units of drift.

ACKNOWLEDGEMENTS

We are very grateful to Paula Rodriguez for her help with insect rearing and to Marco Thali and Joelle Schmid for assistance with the molecular analyses of the pea aphid field samples. We are grateful to Tom Brekke, Meghan Jacobs, Amanda Shaver, Danielle Doi, Evan Tandy, and Katie Mead, who carried out the 2010–2016 screens for *Wolbachia* and *Spiroplasma* infection in *D. neotestacea*. We also thank Ailsa McLean for sharing several *Spiroplasma*-infected aphid lines. This work was supported by a Sinergia grant from the Swiss National Science Foundation (grant No. CRSII3_154396 to CV) and a grant from the US National Science Foundation (grant No. 1144581 to JJ).

AUTHOR CONTRIBUTIONS

H.K., C.H., C.V., J.J., and H.M.H. performed the field sampling; H.K., C.H., and H.M.H. carried out the molecular analysis of the field samples; H.M.H. was responsible for the data analysis and developed the model; H.M.H., H.K., C.H., J.J., and C.V. wrote the paper.

DATA AVAILABILITY STATEMENT

The DNA sequences used in this study are available in Genbank (accession numbers: MG288511–MG288588). The main data set for *D. neotestacea* and *A. pisum* as well as the R function implementing the model are available on Dryad (10.5061/dryad.ch4dp8n).

ORCID

Hugo Mathé-Hubert https://orcid.org/0000-0003-4785-433X
Heidi Kaech https://orcid.org/0000-0002-2149-8050
Corinne Hertaeg https://orcid.org/0000-0002-2730-4042
Christoph Vorburger https://orcid.org/0000-0002-3627-0841

REFERENCES

Anbutsu, H., Lemaitre, B., Harumoto, T., & Fukatsu, T. (2016). Male-killing symbiont damages host's dosage-compensated sex chromosome to induce embryonic apoptosis. *Nature Communications*, *7*, 1–12.

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models. *Biometrika*, *94*, 443–458. https://doi.org/10.1093/biomet/asm017

Balding, D. J., Bishop, M., & Cannings, C. (2007). Population genetics. In D. J. Balding, M. Bishop, & C. Cannings (Eds.), *Handbook of statistical genetics* (pp. 753–780). Chichester, UK: John Wiley & Sons, Ltd.

Ballinger, M. J., Moore, L. D., & Perlman, J. (2018). Evolution and diversity of inherited *Spiroplasma* symbionts in *Myrmica* ants. *Applied and Environmental Microbiology*, *84*, e02299–17.

Ballinger, M. J., & Perlman, S. J. (2017). Generality of toxins in defensive symbiosis: Ribosome-inactivating proteins and defense against parasitic wasps in *Drosophila*. *PLoS Path*, *13*, e1006431. https://doi.org/10.1371/journal.ppat.1006431

Barker, B. (2016). *Aphids in pulse crops*. Saskatchewan Pulse Growers; 8–10.

Brisson, J. A., Nuzhdin, S. V., & Stern, D. L. (2009). Similar patterns of linkage disequilibrium and nucleotide diversity in native and introduced populations of the pea aphid, *Acyrthosiphon pisum*. *BMC Genetics*, *10*, 22. https://doi.org/10.1186/1471-2156-10-22

Brucker, R. M., & Bordenstein, S. R. (2012). Speciation by symbiosis. *Trends in Ecology and Evolution*, *27*, 443–451. https://doi.org/10.1016/j.tree.2012.03.011

Caspi-Fluger, A., Inbar, M., Mozes-Daube, N., Katzir, N., Portnoy, V., Belausov, E., ... Zchori-Fein, E. (2012). Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1791–1796.

Chen, D., & Purcell, A. H. (1997). Occurrence and transmission of facultative endosymbionts in aphids. *Current Microbiology*, *34*, 220–225. https://doi.org/10.1007/s002849900172

Cockburn, S. N., Haselkorn, T. S., Hamilton, P. T., Landzberg, E., Jaenike, J., & Perlman, S. J. (2013). Dynamics of the continent-wide spread of a *Drosophila* defensive symbiont. *Ecology Letters*, *16*, 609–616.

Dedeine, F., Vavre, F., Fleury, F., Loppin, B., Hochberg, M. E., & Bouletreau, M. (2001). Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *Proceedings of the National Academy of Sciences of the USA*, *98*, 6247–6252.

Doremus, M. R., & Oliver, K. M. (2017). Aphid heritable symbiont exploits defensive mutualism. *Applied and Environmental Microbiology*, *83*, 1–15. https://doi.org/10.1128/AEM.03276-16

Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstadter, J., & Hurst, G. D. (2008). The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biology*, *6*, 27. https://doi.org/10.1186/1741-7007-6-27

Everitt, B. S., & Skrondal, A. (2010). *The Cambridge dictionary of statistics*, 4th ed. Edinburgh, UK: Cambridge University Press.

Faria, V. G., Martins, N. E., Magalhães, S., Paulo, T. F., Nolte, V., Schlötterer, C., … Teixeira, L. (2016). Drosophila adaptation to viral infection through defensive symbiont evolution. *PLOS Genetics*, *12*, e1006297. https://doi.org/10.1371/journal.pgen.1006297

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, *78*, 737–756.

Ferrari, J., West, J. A., Via, S., & Godfray, H. C. J. (2012). Population genetic structure and secondary symbionts in host-associated populations of the pea aphid complex. *Evolution*, *66*, 375–390. https://doi.org/10.1111/j.1558-5646.2011.01436.x

Fouedjio, F., & Klump, J. (2019). Exploring prediction uncertainty of spatial data in geostatistical and machine learning approaches. *Environmental Earth Sciences*, *78*, 1–24. https://doi.org/10.1007/s12665-018-8032-z

Frago, E., Mala, M., Weldegergis, B. T., Yang, C., McLean, A., Godfray, H. C. J., … Dicke, M. (2017). Symbionts protect aphids from parasitic wasps by attenuating herbivore-induced plant volatiles. *Nature Communications*, *8*, 1–9. https://doi.org/10.1038/s41467-017-01935-0

Fromont, C., Adair, K. L., & Douglas, A. E. (2019). Correlation and causation between the microbiome, *Wolbachia* and host functional traits in natural populations of drosophilid flies. *Molecular Ecology*, *28*, 1826–1841.

Fukatsu, T., Tsuchida, T., Nikoh, N., & Koga, R. (2001). *Spiroplasma* symbiont of the pea aphid, *Acyrthosiphon pisum* (Insecta: Homoptera). *Applied and Environmental Microbiology*, *67*, 1284–1291. https://doi.org/10.1128/AEM.67.3.1284-1291.2001

Gehrer, L., & Vorburger, C. (2012). Parasitoids as vectors of facultative bacterial endosymbionts in aphids. *Biology Letters*, *8*, 613–615. https://doi.org/10.1098/rsbl.2012.0144

Goryacheva, I., Blekhman, A., Andrianov, B., Romanov, D., & Zakharov, I. (2018). *Spiroplasma* infection in *Harmonia axyridis* – Diversity and multiple infection. *PLoS ONE*, *13*, e0198190. https://doi.org/10.1371/journal.pone.0198190

Haag, C. R., McTaggart, S. J., Didier, A., Little, T. J., & Charlesworth, D. (2009). Nucleotide polymorphism and within-gene recombination in *Daphnia magna* and *D. pulex*, two cyclical parthenogens. *Genetics*, *182*, 313–323.

Hansen, A. K., Vorburger, C., & Moran, N. A. (2012). Genomic basis of endosymbiont-conferred protection against an insect parasitoid. *Genome Research*, *22*, 106–114. https://doi.org/10.1101/gr.125351.111

Hedges, L. M., Brownlie, J. C., O'Neill, S. L., & Johnson, K. N. (2008). *Wolbachia* and virus protection in insects. *Science*, *322*, 702. https://doi.org/10.1126/science.1162418

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518. https://doi.org/10.7717/peerj.5518

Henry, L. M., Maiden, M. C. J., Ferrari, J., & Godfray, H. C. J. (2015). Insect life history and the evolution of bacterial mutualism. *Ecology Letters*, *18*, 516–525. https://doi.org/10.1111/ele.12425

Henry, L. M., Peccoud, J., Simon, J. C., Hadfield, J. D., Maiden, M. J. C., Ferrari, J., & Godfray, H. C. J. (2013). Horizontally transmitted symbionts and host colonization of ecological niches. *Current Biology*, *23*, 1713–1717. https://doi.org/10.1016/j.cub.2013.07.029

Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A., & Werren, J. H. (2008). How many species are infected with *Wolbachia*? – A statistical analysis of current data. *FEMS Microbiology Letters*, *281*, 215–220.

Himler, A. G., Adachi-Hagimori, T., Bergen, J. E., Kozuch, A., Kelly, S. E., Tabashnik, B. E., … Hunter, M. S. (2011). Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science*, *332*, 254–256. https://doi.org/10.1126/science.1199410

Jabot, F., Faure, T., Dumoulin, N., & Albert, C. (2015). *EasyABC: Efficient approximate bayesian computation sampling schemes*. R package version 1.5.

Jaenike, J. (2012). Population genetics of beneficial heritable symbionts. *Trends in Ecology and Evolution*, *27*, 226–232. https://doi.org/10.1016/j.tree.2011.10.005

Jaenike, J., Polak, M., Fiskin, A., Helou, M., & Minhas, M. (2007). Interspecific transmission of endosymbiotic *Spiroplasma* by mites. *Biology Letters*, *3*, 23–25.

Jaenike, J., Stahlhut, J. K., Boelio, L. M., & Unckless, R. L. (2010). Association between *Wolbachia* and *Spiroplasma* within *Drosophila neotestacea*: An emerging symbiotic mutualism? *Molecular Ecology*, *19*, 414–425. https://doi.org/10.1111/j.1365-294X.2009.04448.x

Jaenike, J., Unckless, R., Cockburn, S. N., Boelio, L. M., & Perlman, S. J. (2010). Adaptation via symbiosis: Recent spread of a *Drosophila* defensive symbiont. *Science*, *329*, 212–215. https://doi.org/10.1126/science.1188235

Keightley, P. D., Ness, R. W., Halligan, D. L., & Haddrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics*, *196*, 313–320.

Kimura, M., & Ohta, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, *61*, 763–771.

Kremer, N., Charif, D., Henri, H., Bataille, M., Prévost, G., Kraaijeveld, K., & Vavre, F. (2009). A new case of *Wolbachia* dependence in the genus *Asobara*: Evidence for parthenogenesis induction in *Asobara japonica*. *Heredity (Edinb)*, *103*, 248–256. https://doi.org/10.1038/hdy.2009.63

Lukasik, P., Guo, H., Van Asch, M., Ferrari, J., & Godfray, H. C. J. (2013). Protection against a fungal pathogen conferred by the aphid facultative endosymbionts *Rickettsia* and *Spiroplasma* is expressed in multiple host genotypes and species and is not influenced by co-infection with another symbiont. *Journal of Evolutionary Biology*, *26*, 2654–2661.

Lukasik, P., van Asch, M., Guo, H., Ferrari, J., & Godfray, H. C. J. (2013). Unrelated facultative endosymbionts protect aphids against a fungal pathogen. *Ecology Letters*, *16*, 214–218. https://doi.org/10.1111/ele.12031

Lynch, M., & Deng, H.-W. (1994). Genetic slippage in response to sex. *The American Naturalist*, *144*, 242–261. https://doi.org/10.1086/285673

Margulis, L., & Fester, R. (1991). *Symbiosis as a source of evolutionary innovation: Speciation and morphogenesis*, 1st ed. Amherst, MA: University of Massachusetts.

Mathé-Hubert, H., Kaech, H., Ganesanandamoorthy, P., & Vorburger, C. (2019). Evolutionary costs and benefits of infection with diverse strains of *Spiroplasma* in pea aphids. *Evolution*, *73*, 1–16.

Montllor, C. B., Maxmen, A., & Purcell, A. H. (2002). Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. *Ecological Entomology*, *27*, 189–195. https://doi.org/10.1046/j.1365-2311.2002.00393.x

Moore, W. S. (1995). Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, *49*, 718–726. https://doi.org/10.2307/2410325

Moran, N. A., & Dunbar, H. E. (2006). Sexual acquisition of beneficial symbionts in aphids. *Proceedings of the National Academy of Sciences of the USA*, *103*, 12803–12806.

Niepoth, N., Ellers, J., & Henry, L. M. (2018). Symbiont interactions with non-native hosts limit the formation of new symbioses. *BMC Evolutionary Biology*, *18*, 1–12. https://doi.org/10.1186/s12862-018-1143-z

Oliver, K. M., Degnan, P. H., Burke, G. R., & Moran, N. A. (2010). Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, *55*, 247–266. https://doi.org/10.1146/annurev-ento-112408-085305

Oliver, K. M., Degnan, P. H., Hunter, M. S., & Moran, N. A. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science*, *325*, 992–994. https://doi.org/10.1126/science.1174463

Oliver, K. M., Moran, N. A., & Hunter, M. S. (2006). Costs and benefits of a superinfection of facultative symbionts in aphids. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 1273–1280. https://doi.org/10.1098/rspb.2005.3436

Peccoud, J., Bonhomme, J., Mahéo, F., de la Huerta, M., Cosson, O., & Simon, J. C. (2014). Inheritance patterns of secondary symbionts during sexual reproduction of pea aphid biotypes. *Journal of Insect Science*, *21*, 291–300. https://doi.org/10.1111/1744-7917.12083

Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of the USA*, *106*, 7495–7500.

Pieper, K. E., & Dyer, K. A. (2016). Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. *Journal of Evolutionary Biology*, *29*, 2229–2241. https://doi.org/10.1111/jeb.12948

PLOS Biology Issue Image (2010). The pea aphid *Acyrthosiphon pisum*, an emerging genomic model. *PLOS Biology*, *8*(2), ev08.i02.

Queller, D. C., & Strassmann, J. E. (2016). Problems of multi-species organisms: Endosymbionts to holobionts. *Biology and Philosophy*, *31*, 855–873. https://doi.org/10.1007/s10539-016-9547-x

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Regassa, L. B. (2014). The family *Spiroplasmataceae*. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The prokaryotes: Firmicutes and tenericutes* (pp. 551–567). Berlin, Heidelberg, Germany: Springer.

Rice, J. A. (2006). *Mathematical statistics and data analysis*, 3rd ed. Berkeley, CA: University of California.

Rillig, M. C., Antonovics, J., Caruso, T., Lehmann, A., Powell, J. R., Veresoglou, S. D., & Verbruggen, E. (2015). Interchange of entire communities: Microbial community coalescence. *Trends in Ecology and Evolution*, *30*, 470–476. https://doi.org/10.1016/j.tree.2015.06.004

Rock, D. I., Smith, A. H., Joffe, J., Albertus, A., Wong, N., O'Connor, M., … Russell, J. A. (2017). Context-dependent vertical transmission shapes strong endosymbiont community structure in the pea aphid, *Acyrthosiphon pisum*. *Molecular Ecology*, *27*, 2039–2056.

Russell, J. A., & Moran, N. A. (2006). Costs and benefits of symbiont infection in aphids: Variation among symbionts and across temperatures. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 603–610. https://doi.org/10.1098/rspb.2005.3348

Russell, J. A., Weldon, S., Smith, A. H., Kim, K. L., Hu, Y. I., Łukasik, P., … Oliver, K. M. (2013). Uncovering symbiont-driven genetic diversity across North American pea aphids. *Molecular Ecology*, *22*, 2045–2059. https://doi.org/10.1111/mec.12211

Sanada-Morimura, S., Matsumura, M., & Noda, H. (2013). Male killing caused by a *Spiroplasma* symbiont in the small brown planthopper, *Laodelphax striatellus*. *Journal of Heredity*, *104*, 821–829. https://doi.org/10.1093/jhered/est052

Simon, J.-C., Boutin, S., Tsuchida, T., Koga, R., Le Gallic, J.-F., Frantz, A., … Fukatsu, T. (2011). Facultative symbiont infections affect aphid reproduction. *PLoS ONE*, *6*(7), e21831. https://doi.org/10.1371/journal.pone.0021831

Smith, A. H., Łukasik, P., O'Connor, M. P., Lee, A., Mayo, G., Drott, M. T., … Russell, J. A. (2015). Patterns, causes, and consequences of defensive microbiome dynamics across multiple scales. *Molecular Ecology*, *24*, 1135–1149. https://doi.org/10.1111/mec.13095

Sunnucks, P., & Hales, D. F. (1996). Numerous transposed sequences of mitochondrial cytochrome I-II in aphids of the genus *Sitobion* (*Hemiptera*: *Aphididae*). *Molecular Biology and Evolution*, *13*, 510–524.

Tabata, J., Hattori, Y., Sakamoto, H., Yukuhiro, F., Fujii, T., Kugimiya, S., … ageyama, D. (2011). Male killing and incomplete inheritance of a novel *Spiroplasma* in the moth *Ostrinia zaguliaevi*. *Microbial Ecology*, *61*, 254–263. https://doi.org/10.1007/s00248-010-9799-y

Teixeira, L., Ferreira, A., & Ashburner, M. (2008). The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLOS Biology*, *6*, e1000002. https://doi.org/10.1371/journal.pbio.1000002

Turner, B. M., Sederberg, P. B., & McClelland, J. L. (2014). Bayesian analysis of simulation-based models. *Journal of Mathematical Psychology*, *72*, 191–199. https://doi.org/10.1016/j.jmp.2014.10.001

Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*, 4334–4345. https://doi.org/10.1111/j.1365-294X.2008.03921.x

Vorburger, C. (2006). Temporal dynamics of genotypic diversity reveal strong clonal selection in the aphid *Myzus persicae*. *European Society for Evolutionary Biology*, *19*, 97–107. https://doi.org/10.1111/j.1420-9101.2005.00985.x

Vorburger, C. (2018). Symbiont-conferred resistance to parasitoids in aphids – Challenges for biological control. *Biological Control*, *116*, 17–26. https://doi.org/10.1016/j.biocontrol.2017.02.004

Vorburger, C., & Gouskov, A. (2011). Only helpful when required: A longevity cost of harbouring defensive symbionts. *Journal of Evolutionary Biology*, *24*, 1611–1617. https://doi.org/10.1111/j.1420-9101.2011.02292.x

Weldon, S. R., Strand, M. R., & Oliver, K. M. (2013). Phage loss and the breakdown of a defensive symbiosis in aphids. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 2012–2103. https://doi.org/10.1098/rspb.2012.2103

Werner, T., & Jaenike, J. (2017). *Drosophilids of the Midwest and Northeast*, 1st ed. Rochester, NY: River Campus Libraries, University of Rochester.

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

Wulff, J. A., Buckman, K. A., Wu, K., Heimpel, G. E., & White, J. A. (2013). The endosymbiont *Arsenophonus* is widespread in soybean aphid, *Aphis glycines*, but does not provide protection from parasitoids or a fungal pathogen. *PLoS ONE*, *8*, e62145. https://doi.org/10.1371/journal.pone.0062145

Wulff, J. A., & White, J. A. (2015). The endosymbiont *Arsenophonus* provides a general benefit to soybean aphid (*Hemiptera*: *Aphididae*) regardless of host plant resistance (Rag). *Environmental Entomology*, *44*, 574–581. https://doi.org/10.1093/ee/nvv031

Xie, J., Butler, S., Sanchez, G., & Mateos, M. (2014). Male killing *Spiroplasma* protects *Drosophila melanogaster* against two parasitoid wasps. *Heredity (Edinb)*, *112*, 399–408. https://doi.org/10.1038/hdy.2013.118

Zchori-Fein, E., Lahav, T., & Freilich, S. (2014). Variations in the identity and complexity of endosymbiont combinations in whitefly hosts. *Frontiers in Microbiology*, *5*, 1–8. https://doi.org/10.3389/fmicb.2014.00310

Zchori-Fein, E., & Perlman, S. J. (2004). Distribution of the bacterial symbiont *Cardinium* in arthropods. *Molecular Ecology*, *13*, 2009–2016. https://doi.org/10.1111/j.1365-294X.2004.02203.x

Zytynska, S. E., & Weisser, W. W. (2016). The natural occurrence of secondary bacterial symbionts in aphids. *Ecological Entomology*, *41*, 13–26. https://doi.org/10.1111/een.12281

SUPPORTING INFORMATION

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

How to cite this article: Mathé-Hubert H, Kaech H, Hertaeg C, Jaenike J, Vorburger C. Nonrandom associations of maternally transmitted symbionts in insects: The roles of drift versus biased cotransmission and selection. *Mol Ecol*. 2019;28:5330–5346. https://doi.org/10.1111/mec.15206