

Uncovering symbiont-driven genetic diversity across North American pea aphids

JACOB A. RUSSELL,* STEPHANIE WELDON,† ANDREW H. SMITH,* KYUNGSUN L. KIM,† YI HU,* PIOTR ŁUKASIK,* STEVEN DOLL,* IOANNIS ANASTOPOULOS,* MATTHEW NOVIN* and KERRY M. OLIVER†

*Department of Biology, Drexel University, 3245 Chestnut St., Philadelphia, PA 19104, USA, †Department of Entomology, University of Georgia, 413 Biological Sciences Building, Athens, GA 30602, USA

Abstract

Heritable genetic variation is required for evolution, and while typically encoded within nuclear and organellar genomes, several groups of invertebrates harbour heritable microbes serving as additional sources of genetic variation. Hailing from the symbiont-rich insect order Hemiptera, pea aphids (*Acyrtosiphon pisum*) possess several heritable symbionts with roles in host plant utilization, thermotolerance and protection against natural enemies. As pea aphids vary in the numbers and types of harboured symbionts, these bacteria provide heritable and functionally important variation within field populations. In this study, we quantified the cytoplasmically inherited genetic variation contributed by symbionts within North American pea aphids. Through the use of Denaturing Gradient Gel Electrophoresis (DGGE) and 454 amplicon pyrosequencing of 16S rRNA genes, we explored the diversity of bacteria harboured by pea aphids from five populations, spanning three locations and three host plants. We also characterized strain variation by analysing 16S rRNA, housekeeping and symbiont-associated bacteriophage genes. Our results identified eight species of facultative symbionts, which often varied in frequency between locations and host plants. We detected 28 cytoplasmic genotypes across 318 surveyed aphids, considering only the various combinations of secondary symbiont species infecting single hosts. Yet the detection of multiple *Regiella insecticola*, *Hamiltonella defensa* and *Rickettsia* strains, and diverse bacteriophage genotypes from *H. defensa*, suggest even greater diversity. Combined, these findings reveal that heritable bacteria contribute substantially to genetic variation in *A. pisum*. Given the costs and benefits of these symbionts, it is likely that fluctuating selective forces play a role in the maintenance of this diversity.

Keywords: biodiversity, co-infection, endosymbiont, strain diversity, *Wolbachia*

Received 20 September 2012; revision received 4 December 2012; accepted 11 December 2012

Introduction

Intimate, prolonged associations among two or more different organisms—symbioses—are arguably ubiquitous among animals, which harbour microbial consortia comprising numerous bacteria, fungi, protists and viruses. Symbiotic interactions can vary within host species, with individuals differing in their array of resident microbial species (Lee *et al.* 2008; Ishak *et al.* 2011).

Such variation may be of functional importance given the known roles of microbial symbionts in host defense and nutrition (Douglas 2009; Oliver *et al.* 2009a). The most diverse animal groups, including arthropods and nematodes, are often infected with heritable microbes, which are stably transmitted from parent to offspring (Moran & Baumann 2000; Werren *et al.* 2008). Theory predicts, and experimental studies confirm, that heritable symbionts can invade and persist in host populations by providing net fitness benefits to infected hosts relative to uninfected hosts (Bull 1983; Oliver *et al.* 2008). As such, symbionts can serve as an additional

Correspondence: Jacob Russell, Fax: 215 895 1273; E-mail: jar337@drexel.edu

source of heritable genetic variation that can be acted upon by natural selection.

Molecular surveys indicate that the majority of insect species are infected with heritable symbionts (Hilgenboecker *et al.* 2008), with a diverse array of taxa having colonized a broad suite of insects (Duron *et al.* 2008), and with several insects harbouring diverse menageries of heritable symbionts (Sandström *et al.* 2001; Moran *et al.* 2008; Oliver *et al.* 2010; Skaljic *et al.* 2010; Toju & Fukatsu 2011). In most cases, the effects of infection on host biology remain unknown, but aphids have emerged as an important insect model for identifying phenotypic effects. These insects feed exclusively on a nitrogen-poor phloem diet and almost all require the nutritional bacterial symbiont *Buchnera aphidicola* (Douglas 1998). Aphids may also harbour one or more additional heritable bacteria, called 'secondary symbionts'. These can enhance survival and reproduction under particular conditions, although they are not generally required, and can even be costly in some contexts (Oliver *et al.* 2008, 2010; Simon *et al.* 2011; Vorburger & Gouskov 2011).

The pea aphid (*Acyrtosiphon pisum*), in particular, possesses a diverse collection of seven known secondary symbionts (Oliver *et al.* 2010; Tsuchida *et al.* 2010), which are found at intermediate frequencies in natural populations (Montllor *et al.* 2002; Tsuchida *et al.* 2002; Darby *et al.* 2003; Haynes *et al.* 2003; Leonardo & Muir 2003; Simon *et al.* 2003; Ferrari *et al.* 2004, 2012; Oliver *et al.* 2006; Frantz *et al.* 2009). These bacteria are transmitted with high fidelity from mother to offspring (Chen & Purcell 1997; Darby & Douglas 2003; Moran & Dunbar 2006), although they occasionally move horizontally within and among host lineages (Russell *et al.* 2003; Moran & Dunbar 2006; Oliver *et al.* 2008; Gehrer & Vorburger 2012). From a functional perspective, pea aphid secondary symbionts have been shown to mediate important ecological interactions (Oliver *et al.* 2010): *Hamiltonella defensa* infected with APSE bacteriophage (Oliver *et al.* 2003, 2005, 2009b), and possibly the X-type symbiont (Guay *et al.* 2009), defends against parasitic wasps; *Regiella insecticola*, *Rickettsia*, *Rickettsiella* and some strains of *Spiroplasma* protect against fungal pathogens (Scarborough *et al.* 2005; Łukasik *et al.* 2012); *Serratia symbiotica* and *Rickettsia* (and possibly *H. defensa* (Russell & Moran 2006)) provide protection against the damaging effects of high temperatures (Chen *et al.* 2000; Montllor *et al.* 2002); and some *R. insecticola* strains influence host plant utilization (Leonardo 2004; Tsuchida *et al.* 2004; Ferrari *et al.* 2007). Other effects include the alteration of aphid body colour by *Rickettsiella* sp. (Tsuchida *et al.* 2010), which may influence interactions with natural enemies, and male-killing by *Spiroplasma* symbionts (Simon *et al.* 2011). Costs of symbiont infection have also been identified for some pea aphid

symbionts (Chen *et al.* 2000; Oliver *et al.* 2008; Simon *et al.* 2011), suggesting that their intermediate prevalence in natural populations may arise partially due to balancing selection.

In addition to the known range of symbiont lineages, pea aphids appear to harbour multiple strains of some symbiont species (Degnan & Moran 2008; Ferrari *et al.* 2012), suggesting an additional source of heritable cytoplasmic variation. Importantly, distinct APSE phage variants harboured by different *H. defensa* (Degnan & Moran 2008; Oliver *et al.* 2009b) confer different levels of defense against parasitoids (Oliver *et al.* 2005, 2009b). Different *R. insecticola* and *S. symbiotica* strains also differentially impact host phenotypes and fitness, illustrating that costs and benefits may commonly vary within symbiont species (Russell & Moran 2006; Ferrari *et al.* 2007). Aside from these findings, however, our understanding of strain diversity in pea aphids remains limited.

An additional source of symbiont-driven cytoplasmic variation results from infection by various combinations of secondary symbionts (i.e. superinfection). Indeed, surveys of European populations indicated that *A. pisum* can be infected with up to four secondary symbionts per individual [average = 1.4 (Ferrari *et al.* 2012)]. Yet few studies of pea aphids have screened for the full range of known secondary symbionts, while few have used methods capable of detecting unexpected lineages (but see Haynes *et al.* 2003), limiting our knowledge of the frequency and variety of natural superinfections.

To elucidate the diversity of secondary symbionts and their contribution toward cytoplasmically inherited variation, we studied pea aphids from five populations in North America, a region with surprisingly little exploration of symbiont diversity. While studies from Europe and Japan targeted multiple populations (Tsuchida *et al.* 2002; Darby *et al.* 2003; Haynes *et al.* 2003; Simon *et al.* 2003; Ferrari *et al.* 2012), previous N. American investigations were geographically limited and focused on a small set of symbionts through diagnostic PCR (Leonardo & Muir 2003; Oliver *et al.* 2006). Here we used Denaturing Gradient Gel Electrophoresis (DGGE) of 16S rRNA PCR products amplified with universal primers, allowing for the identification of not just the known secondary symbionts, but other bacteria with possible symbiotic status. In addition, we also targeted subsets of collected aphids with 454 amplicon pyrosequencing of 16S rRNA genes (amplified with universal bacterial primers) and diagnostic PCR, providing further opportunities to explore pea aphid-associated bacteria. Analyses of sequences generated through pyrosequencing also allowed us to survey for potential 16S rRNA polymorphisms, while Sanger sequencing of 16S rRNA (from *H. defensa*, *R. insecticola* and *Rickettsia*) and two additional genes (from *H. defensa* and APSE) enabled initiation of a broader study on sym-

biont strain diversity. Our findings reveal that pea aphid cytoplasmic genotypes are highly diverse due to the presence of multiple symbiont strains, variable infection levels by seven common symbiont species, and the frequent occurrence of superinfection by multiple symbiont species.

Materials and methods

Collections and extractions

Pea aphids were collected from alfalfa (*Medicago sativa*), red clover (*Trifolium pratense*) and hairy vetch (*Vicia villosa*) after beating plants over a plastic tray. While the targeted populations were expected to be cyclically parthenogenetic (based on latitude), we attempted to minimize resampling of the same clones by collecting single parthenogenetic aphids at 20+ meter intervals. Collections from Pennsylvania (PA) alfalfa targeted three fields in Montgomery County, with collection dates ranging from April 23 through August 25, 2010. We also collected from two fields of clover and one field of hairy vetch in PA, with sampling dates ranging from May 7 through June 15, 2010. Utah (UT) collections targeted aphids from alfalfa on August 18, 2011, while collecting in Wisconsin (WI) targeted aphids from alfalfa on September 14, 2011. Each of the above populations was used in our assessments of symbiont frequencies. For our sequence-based studies, we also included alfalfa-derived aphids collected from New York state (NY) and Pennsylvania in 2011. Aphids from PA and NY were immediately preserved in ethanol for subsequent screening, while those from UT and WI were brought to the laboratory for rearing and the eventual screening of first generation laboratory aphids.

DNA was extracted from all PA and NY aphids using a previously described protocol (Russell *et al.* 2003). For aphids from WI and UT, field-collected nymphs were grown to adulthood on *Vicia faba* in the laboratory and DNA extractions were conducted on eight 3rd instar first generation offspring per clonal line using E.Z.N.A. tissue DNA kit (Omega Bio-Tek, Norcross, GA, USA) following the manufacturer's protocol. Template quality was then assessed, for all aphids, through amplification of bacterial DNA with universal PCR primers (described in more detail below).

Denaturing gradient gel electrophoresis

Using whole genomic DNA extracted from aphids and their symbionts, we PCR amplified a 16S rRNA gene fragment of approximately 190 bp (corresponding to *E. coli* 16S positions 340 - 533) using universal primers 356F (with a 40 bp 'GC-clamp') and 517R (Myers *et al.*

1985; Bano & Hollibaugh 2002) (see Table S2 and Appendix S1 from the Supporting information for primer sequences, PCR cocktail recipes, and cycling conditions).

Prior to DGGE analysis, PCR product amplification was verified on 1% agarose gels stained with Biotum GelRed nucleic acid dye (Biotum, Hayward, CA, USA). DGGE was then performed using a CBS Scientific DGGE system (Del Mar, CA, USA). Briefly, about 300 ng of PCR template for each sample was loaded on a 6.5% polyacrylamide gel containing a continuous 40–65% denaturing gradient (100% denaturant is 7 M urea and 40% formamide). Gels were run at 70 V for approximately 16.5 h in 1× TAE buffer at 60 °C.

Band excision and sequencing (see Appendix S1, Supporting information for detail) revealed that all seven detected symbionts produced a single diagnostic band, with the exception of *H. defensa*, which produced either one (as seen in the standard on Fig. S1, Supporting information) or four bands (as seen in the two lanes to the left of the standard). Sequencing of these four bands revealed only one variable nucleotide within the targeted amplicon (Table S1, Supporting information).

After associating bands with particular symbionts, a 'standard' of common *A. pisum* symbionts was created by mixing genomic DNA in equal concentrations (50 ng) from single extracted bands. This mixture was then used as template for PCR using 356F and 517R primers. PCR results from this standard sample were then run on all DGGE gels as a reference to identify common symbionts (Fig. S1). Bands that were faint and those that could not be assigned to a known symbiont were excised from gels, re-amplified, purified and sequenced (see Appendix S1, Supporting information for more detail). These partial 16S rRNA sequences were analysed for DNA sequence similarity using the BLASTN algorithm against the 'non-redundant' GenBank (NCBI) nucleotide database.

Diagnostic PCR was conducted (see Table S2, Supporting information for primers and protocols) to confirm all positive DGGE results, while negative DGGE results (suggesting no secondary symbionts) were reassessed with universal 16S-23S rRNA primers (559F and 35R, Table S2, Supporting information) that amplify most bacteria, but not *Buchnera*. In addition to surveys for secondary symbionts, screening for APSE bacteriophages was performed with primers targeting the P2 and P28 genes (Table S2, Supporting information).

Statistics on symbiont frequencies

Symbiont frequencies were estimated separately for each of the five studied pea aphid populations. Subsequently, two-tailed Fisher's exact tests were performed to assess the significance of frequency differences between host

race populations from the same region (Pennsylvania) or populations from different locations (Utah, Wisconsin, and Pennsylvania alfalfa populations). To assess whether symbiont frequencies changed over time in PA alfalfa populations, collection periods were binned into 4-week 'quarters' (spanning April 23–August 25) due to insufficient sample sizes on each specific collection date. Two-tailed Fisher's exact tests were then performed to assess whether bacterial frequencies changed between adjacent time points.

We compared the numbers of symbionts found, on average, within aphid hosts between populations using *t*-tests. Bonferroni corrections were applied to control for Type I error. To test whether particular single and superinfection types were found more or less frequently than expected based on chance, we generated null expectations for each observed infection type. To do this, we first computed the expected numbers of aphids with the given infection type for each population. We then added expected numbers from all five populations together, using two-tailed Fisher's exact tests to compare this value to the observed number of aphids with the same infection type. A similar approach was taken to assess deviations from null expectations for all observed pairwise combinations of symbionts (regardless of whether particular pairs were found in double, triple or quadruple infections), and to also assess whether the incidence of superinfection differed from null expectations (see Appendix S1, Supporting information for more detail).

For all Fisher's exact test statistics, expected values were rounded to the nearest integer, while Bonferroni corrections were applied to control for Type I error arising due to multiple comparisons. For statistics on superinfection, comparisons were only made if the observed or expected number of infections was ≥ 10 .

454 pyrosequencing and the identification of bacterial associates of aphids

To provide an independent means of identifying microbes from aphids, we performed 454 amplicon pyrosequencing of bacterial 16S rRNA genes using DNA pooled from single aphids. DNA from single aphid extractions was first quantified using a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA). After quantification, DNA samples from 18 to 25 aphids (average = 21) were normalized and combined to create DNA pools. Ten pools were created in total: two each from alfalfa-collected aphids from PA, UT and WI, and two each from aphids derived from clover and hairy vetch in PA.

Samples were then sent to the Research and Testing Laboratory (Lubbock, TX, USA) for bacterial tag-encoded

FLX 454 pyrosequencing of the V1-V3 variable regions of the 16S rRNA gene with primers Gray28F (5'-GAGTTT GATCNTGGCTCAG-3') and Gray519R (5'-GTNTTACN GCGGCKGCTG-3').

Raw sequences were analysed using the QIIME pipeline version 1.5 (Caporaso *et al.* 2010). We used the default quality control parameters to remove all low-quality or ambiguous reads based on a minimum sequence quality score of 25 and a minimum read length of 200 bp. All sequences with mismatches in barcode or primer regions were also discarded. After quality-filtered sequences were denoised by the QIIME denoiser, reads were grouped into 97% OTUs using *uc-lust* (Edgar 2010). Chloroplast sequences were identified and removed after classification on the Ribosomal Database Project (RDP) website (Cole *et al.* 2005). We initially removed chimeras with the chimera slayer algorithm on QIIME; however, we found that a large cluster of sequences from *S. symbiotica* were mistakenly classified as chimeric. As we aimed to include these sequences in our analyses, we eliminated the chimera-checking step, meaning that some of the other analysed sequences may indeed have been chimeric.

After grouping sequences into 67 distinct 97% OTUs (excluding chloroplasts), a file with the longest sequence from each OTU was uploaded to RDP where sequences were classified to the lowest taxonomic level with 80% bootstrap support. SeqMatch searches were performed using RDP, and an alignment was performed on top SeqMatch hits along with representative 454 sequences from each OTU (on the RDP website). Using this alignment, we generated a maximum likelihood phylogeny using the program GARLI (Zwickl 2006) on the CIPRES portal (Miller *et al.* 2012). The phylogenetic analysis employed a GTR + G + I model of nucleotide substitution, and model parameters were estimated during the run.

The tree with the highest likelihood score was inspected to identify the closest relatives of each representative 454 sequence for all OTUs. Multiple OTUs were found to cluster tightly with *Buchnera*, *S. symbiotica* and *H. defensa*, likely due to OTU inflation resulting from sequencing error. We subsequently coalesced OTUs from these respective symbionts, binning sequence reads into single groups. After this step, the proportions of sequence reads from each OTU or symbiotic group were estimated for all 10 samples/DNA pools, giving insight into symbiont diversity, distributions and abundance (see Table S3, Supporting information for details).

Identifying 16S rRNA polymorphisms

All 454 sequences from OTUs clustering with known secondary symbionts were separately uploaded to RDP for

sequence alignment. Alignments were then inspected to identify candidate sequence polymorphisms among each of the seven known secondary symbionts. To avoid inclusion of likely sequencing errors, we ignored sequence variants containing indels (typically occurring within homopolymer tracts) and those at low frequency. While we did not apply a strict, quantitative means to eliminate rare variants, all of those included in our analysis as candidate polymorphisms exceeded 20% representation in at least one of the 10 sequence libraries, reflecting the conservatism of our approach.

To further increase the stringency of our search for candidate polymorphisms, we focused on those found across multiple sequence libraries/DNA pools and/or those from haplotypes that were identical or nearly identical to previously generated 16S rRNA sequences. After applying these criteria, we quantified the proportional representation of different sequence haplotypes from each of our libraries, giving insight into the distribution of different symbiont strains across pea aphid populations.

To aid our search for multiple symbiont strains in pea aphid populations, we also performed a series of diagnostic PCRs targeting portions of three distinct loci: (i) the 5' end of the 16S rRNA gene for *R. insecticola*, *Rickettsia* and *H. defensa*; (ii) the *recJ* gene of *H. defensa*; and (iii) the P3 gene of the APSE bacteriophage from *H. defensa* (a portion of the adjacent P2 gene was also sequenced, although this was not included in phylogenetic analyses). The latter two loci were chosen due to their modest-to-high levels of variability among *H. defensa* and APSE (see Table S2, Supporting information for primers and reaction conditions). Amplified products were purified using an Exo-AP protocol. First, 1 µL of Antarctic phosphatase, 1 µL of Antarctic phosphatase buffer and 0.6 µL of *E. coli* exonuclease I (New England Biolabs, Ipswich, MA, USA) were added to 20 µL PCR product. Reactions were then held at 37 °C for 35 min, followed by 80 °C for 20 min. Purified products were then sent to the UPenn Sequencing Center for single direction Sanger sequencing (16S rRNA and *recJ*) or sequencing with multiple primers (P2 and P3 genes).

Sequences were assembled and manually edited in SEQUENCHER version 4.2 (Genecodes 2003) or CODONCODE ALIGNER version 4.0.2 (CodonCode Corporation 2012) before use in BLASTN searches. Top hits and outgroups were downloaded and aligned with Sanger and 454 sequences from secondary symbionts using the Muscle algorithm (Edgar 2004) in the program SEAVIEW (Gouy *et al.* 2010). Maximum likelihood phylogenies were generated with these alignments using GARLI on the CIPRES Portal (as described above). The default settings of PHYML 3.0 (Guindon *et al.* 2010) in the SEAVIEW package were applied for bootstrap analyses (100 replicates).

Results

Symbiont screening

Denaturing Gradient Gel Electrophoresis screening revealed that North American *A. pisum* are infected with a diverse array of secondary symbionts (Fig. 1; Table S4, Supporting information), and the majority of aphids (mean = 74.2%) were infected with at least one such microbe. All *H. defensa*-infected aphids harboured APSE bacteriophage ($n = 103$) according to diagnostic PCR, indicating that phage-free *H. defensa*-infected *A. pisum* are uncommon in nature.

Diverse symbiont assemblages were evident at the population level at least five heritable symbionts were found in every population, while all seven originally described symbionts were present in two of alfalfa populations. We also detected *Wolbachia* infecting a single PA clone, bringing the total number of secondary symbionts reported in pea aphids to eight. The *Wolbachia*-infected clone was derived from a subset of samples not included in our quantitative analyses because they were not screened immediately following collection (i.e. field-collected aphids or their first generation progeny) and had potentially lost symbionts present in the field (results for these aphids are in Table S4, Supporting information). This clone had been laboratory reared in the absence of parasitoids for several months prior to screening, indicating that *Wolbachia* was vertically transmitted over several generations and that detection did not result from contamination caused by *Wolbachia*-infected parasitoids. Unfortunately, the infection was lost shortly after discovery, preventing further study. The recovered sequence appeared to cluster within Supergroup A and was 100% identical to other *Wolbachia* strains described from insects such as whiteflies, scale insects, parasitoid wasps and other aphids (see Table S1, Supporting information for recovered *Wolbachia* sequence identified through DGGE).

Overall, infection frequencies for the seven targeted symbionts varied from 0% to 67.7%, with differences occurring among populations (see Table 1 for significant results and Table S5 Supporting information for all population comparison statistics). For instance, *S. symbiotica* was found in 47.4% of $n = 38$ surveyed aphids from WI alfalfa, an infection level that significantly exceeded that observed among pea aphids from PA alfalfa (6.0% of 116 aphids). Aphids from this latter population also harboured fewer *S. symbiotica* compared to those from vetch in PA (30%, $n = 50$). *Rickettsiella* symbionts were significantly more prevalent within the WI alfalfa population (47.4% of 38 aphids) compared to both of the remaining alfalfa populations (17.2%,

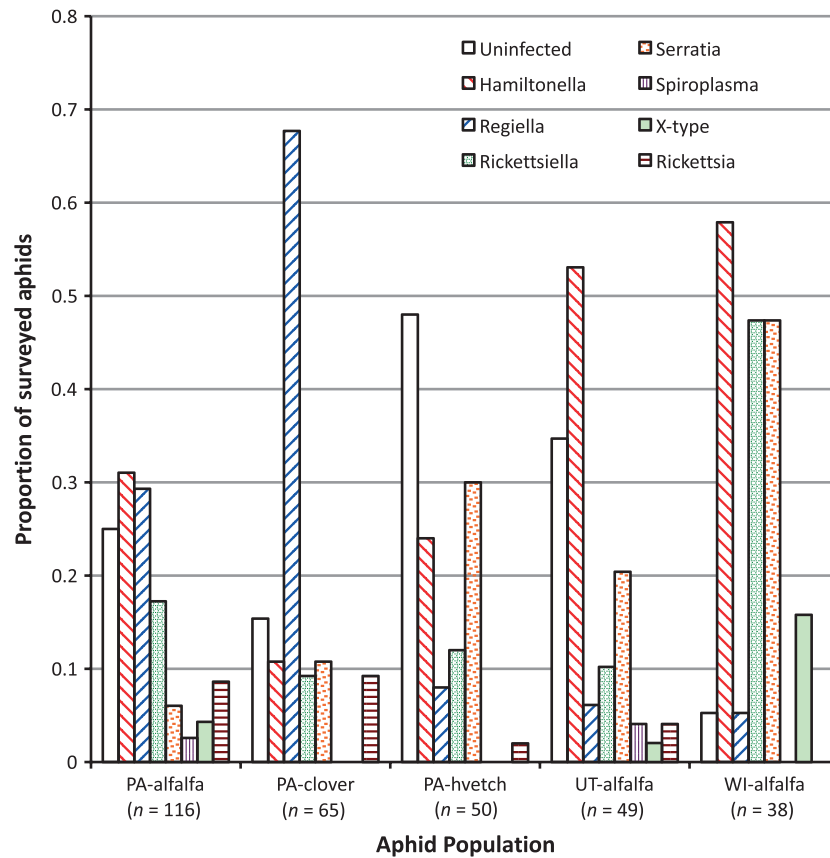


Fig. 1 Frequencies of secondary symbionts across pea aphids from five populations. Proportions of aphids harbouring each of the seven major secondary symbionts are graphed on the y-axis along with the proportions of 'uninfected' aphids free of secondary symbionts. Sample sizes for numbers of aphids screened across five populations (three geographic locations and three host races) are provided below population names.

$n = 116$ for PA alfalfa; 10.2%, $n = 49$ for UT alfalfa). In addition, *R. insecticola* frequencies were significantly lower within PA vetch (8%, $n = 50$) and PA alfalfa (29.3%, $n = 116$) populations when compared with the population from PA clover (67.7%, $n = 65$). Finally, *H. defensa* frequencies differed between PA clover (10.8%, $n = 65$) and PA alfalfa (31.0%, $n = 116$) populations, with significantly higher levels in the latter.

As PA populations from different host plants were collected over different time spans, we performed a more controlled series of statistical comparisons, assessing frequency differences during a time when collections were made for all three host plant populations (weeks 3–9). Most of the significant differences seen for the entire data set remained significant in this more controlled data set, revealing robustness of host plant differences (Table S5, Supporting information).

When comparing frequencies of symbionts over time from our most thoroughly sampled population (PA alfalfa), we observed only marginally significant differences ($0.1 > P > 0.05$) before applying the Bonferroni correction (Table S6, Supporting information). Given the

Table 1 Aphid populations showing significantly different symbiont frequencies after Bonferroni correction

Population 1*	Population 2	Infection type	Fisher's exact test two-tailed P-values
PA vetch	PA clover	uninfected	0.0002
PA alfalfa	PA clover	<i>Hamiltonella</i>	0.0001
PA clover	PA alfalfa	<i>Regiella</i>	0.0001
PA clover	PA vetch	<i>Regiella</i>	0.0001
PA alfalfa	UT alfalfa	<i>Regiella</i>	0.0009
WI alfalfa	UT alfalfa	<i>Rickettsiella</i>	0.0002
WI alfalfa	PA alfalfa	<i>Rickettsiella</i>	0.0004
WI alfalfa	PA alfalfa	<i>Serratia</i>	0.0001
PA vetch	PA alfalfa	<i>Serratia</i>	0.0001

*The population with the higher average frequency is listed in the first column.

sample sizes for each analysed quarter ($n = 25$ – 33); however, statistical power was probably too low to detect low-to-modest frequency shifts.

Table 2 Numbers of secondary symbionts per aphid

Population	0 SS*	1 SS	2 SS	3 SS	4 SS	Average # symbionts/ aphid†
PA alfalfa	29	62	22	3	0	0.991 ^{A,a}
PA clover	10	44	8	2	1	1.077 ^A
PA vetch	24	17	6	3	0	0.760 ^A
UT alfalfa	17	20	7	5	0	1.000 ^a
WI alfalfa	2	15	12	9	0	1.737 ^b
Total	82	158	55	22	1	1.063

*'SS', secondary symbiont; numbers listed in columns 2–6 are the numbers of aphids harbouring the given number of secondary symbionts.

†Capital 'A' symbols reveal no significant differences among PA populations, after Bonferroni corrections, while lowercase 'b' vs. 'a' symbols reveal significantly higher numbers per aphid for WI vs. PA and UT alfalfa populations.

Table 3 Rare and common superinfection types across 318 pea aphids

Symbionts	Observed #	Expected #	P-value*
<i>Regiella</i> only	66	47.3	0.0615
<i>Hamiltonella</i> - <i>Rickettsiella</i> - <i>Serratia</i>	15	5.0	0.0383
<i>Hamiltonella</i> - <i>Regiella</i>	2	12.3	0.0120

*Significant or marginally significant according to two-tailed Fisher's exact test, but not after Bonferroni correction.

Across most populations, superinfections with multiple secondary symbionts were common, ranging from 16.9% in PA clover to 55.3% in WI alfalfa. Of the 318 screened pea aphids, 78 harboured more than one secondary symbiont, with 55 double infections, 22 triple infections and one quadruple infection (vs. 158 single infections) (Table 2; Table S7, Supporting information). Overall, we detected significantly fewer multiple infections than expected by chance given the known infection frequencies (139 expected vs. 78 observed multiple infections; $n = 318$ aphids; two-tailed Fisher's exact test $P = 0.0001$). However, after removing all aphids with *Regiella* symbionts from the analysis (*Regiella* were often found in single infections; see Table 3), this difference became nonsignificant (53 expected vs. 57 observed multiple infections; $n = 231$ aphids; two-tailed Fisher's exact test $P = 0.7432$).

Our assessment of symbiont presence/absence identified 28 infection types, including aphids uninfected with secondary symbionts, those with one of the seven expected single infections, and aphids with 20 additional types of superinfection. Among these, we found

a small number that deviated from random expectations estimated from symbiont frequencies (Table 3; Table S8, Supporting information). When examining pairwise combinations of symbionts (which assessed co-occurrence of symbiont pairs across all double, triple and quadruple infections; Table S9, Supporting information), *S. symbiotica* and *Rickettsiella* appeared more commonly within the same hosts than expected by chance (10 expected vs. 24 observed instances; $n = 318$ aphids; Fisher's exact test $P = 0.0207$), while *H. defensa* and *R. insecticola* were found in fewer aphids than expected (28 expected vs. seven observed instances; $n = 318$ aphids; Fisher's exact test $P = 0.0004$). It should be noted, however, that none of the deviations seen for superinfections remained significant after Bonferroni correction, indicating the need for further sampling to re-assess these specific patterns of superinfection.

Detection of symbionts through 454 sequencing

A total of 55 378 16S rRNA sequence reads were obtained from the 10 pooled samples of aphid DNA, with a range of 2838–16 637 reads, and a median of 4199 reads per sample. A maximum likelihood phylogeny of top SeqMatch hits and representative sequences from our data set allowed us to characterize each of the 67 (97%) OTUs from our data set, and to assign them to taxa (and symbiotic groups) based on their closest relatives.

In total, *Buchnera* sequences comprised an average of 88.4% of the sequence reads, being found across all pools. *Serratia symbiotica* was the next most abundant, with representation in all five populations and an average sequence read abundance of 4.3%. The X-type, *Rickettsia*, *H. defensa* and *R. insecticola* were next in read abundance, ranging from 1.2 to 1.7% of sequence reads, on average, across all samples. These symbionts were detected in two, four, five and three populations, respectively. Following these symbionts (and several other bacteria not known to be symbiotic) were *Rickettsiella* and *Spiroplasma*, which were each found in one population, with .06 and .03% average representation (see Table S3 and Fig. S2, Supporting information for details).

Overall, 454 sequencing identified the same heritable lineages detected by DGGE, but there was little quantitative agreement between symbiont frequency data (as ascertained with DGGE) and read numbers obtained through 454 sequencing. For instance, the X-type and *Rickettsia* symbionts ranked second and third in average sequence read abundance for secondary symbionts, despite never exceeding fourth in terms of symbiont frequency across the five surveyed populations. In contrast, *Rickettsiella* symbionts were not detected in four of

the surveyed populations with 454 surveys, in spite of their presence in all five populations (and occasionally high prevalence) according to DGGE. Although the reasons for these discrepancies are currently unclear, they could result from methodological bias or differences in symbiont density (e.g. high X-type infection densities, low densities of *Rickettsiella*), as the 454 data are also influenced by the abundance of bacteria within the sampled hosts.

Aside from the aforementioned symbionts, 454 sequencing detected other bacteria not known to be heritable in insects, although none exceeded an average of 0.6% of 454 sequence reads. Those distributed across multiple populations were related to a *Pantoea* species from another aphid, to *Ralstonia* from human mouths, to an *Acinetobacter* species, to an *Enterobacter* species previously found in a whale and to a *Propionibacterium* species from humans. It is possible that some of these (and other detected) bacteria could be contaminants, but sequencing of DGGE bands did identify ephemeral bacteria related to *Pantoea*, *Acinetobacter* and *Enterobacter* (Table S2, Supporting information). Although heritability cannot be ruled out, it is more likely that these microbes engage in opportunistic associations with aphids—perhaps as gut associates, pathogens—or that they represent contaminants from soil, plants or human handling.

Strain variation within symbiont species

Our 454 sequence analyses revealed clear evidence for sequence variation in the 16S rRNA genes of *R. insecticola* and *H. defensa*: each exhibited two polymorphic sites, resulting in two suspected haplotypes for *R. insecticola* and three for *H. defensa*. All but one of these haplotypes was recovered with direct, Sanger sequencing of diagnostic PCR products. Among the Sanger sequences, only one showed evidence for multiple peaks at a suspected polymorphic site (although most *H. defensa* sequences yielded multiple peaks at other positions). In this case, the cause was likely due to the presence of multiple *H. defensa* strains, as the variable sites showed no differences between operons of a sequenced *H. defensa* genome (Accession #CP001277; analysis not shown). Conversely, we found no evidence for common sequence variation among *Spiroplasma*, *Rickettsiella* or X-type symbionts (based on 454 sequencing of 16S rRNA genes), although the former two had sparse coverage in 454 sequencing ($n = 15$ and $n = 25$ sequence reads, respectively). Apparent polymorphisms in *S. symbiotica* were probably caused by inter-operon differences, as suggested by the consistent presence of multiple chromatogram peaks at the suspected polymorphic sites (Sanger sequencing data not shown).

While we could not confirm the likely polymorphisms for *Rickettsia* through Sanger sequencing, the nucleotide sequences of two haplotypes identified through 454 sequencing were either highly similar (i.e. 397/398 nucleotides identical to multiple *Rickettsia bellii* strains) or identical to those from previously sequenced strains. Thus, the observed sequence variation in *Rickettsia* is likely not an artefact.

To illustrate the presence of the different sequence variants detected from our 454 pyrosequencing, we graphed the frequency of each sequence haplotype across pea aphid populations for *H. defensa*, *R. insecticola* and *Rickettsia* (Fig. S3, Supporting information). Results suggest that multiple *Rickettsia* strains existed in pea aphids from PA clover populations. Similarly, multiple strains of *R. insecticola* were found in aphids from PA alfalfa. In contrast, all but one of the five studied populations appeared to harbour multiple *H. defensa* strains, with the WI and UT alfalfa populations potentially harbouring three 16S rRNA variants.

Phylogenetic analyses of Sanger and 454 sequencing results provided further support for the aforementioned trends, suggesting the presence of multiple *H. defensa*, *R. insecticola* and *Rickettsia* strains within pea aphids. First, 16S rRNA phylogenies combining our sequences, plus additional sequences from GenBank, revealed multiple lineages of each symbiont associated with pea aphids, suggesting independent acquisitions of different symbiont strains (Fig. 2a; Figs S4–S5, Supporting information). Interestingly, the aphids *Sitobion miscanthi* and *S. avenae* also appear to harbour multiple symbiont strains with independent origins, based on our analyses of unpublished GenBank sequence data.

Analyses of additional genes encoded by *H. defensa* and their APSE phage corroborated trends seen for 16S rRNA genes. Specifically, phylogenetic analyses of the *H. defensa* gene *recJ* revealed five distinct variants in pea aphids (four found here), including three newly discovered in the course of this study (Fig. 2b). Up to three variants were found to coexist within individual populations, although just one variant was detected from UT aphids. Analyses of the P3 gene encoded by APSE phage showed even higher levels of diversity, with a total of nine allelic lineages from pea aphids alone (Fig. 2c). Four of these lineages were newly identified in our study, and a total of seven P3 variants were found in pea aphids from populations in NY, PA, UT and WI, with two to four in each.

All three *H. defensa*-associated loci (16S rRNA, *recJ* and P3) were sequenced from 34 clonal aphid lines. From this data set, we identified seven multi-locus genotypes, including some probably arising via recombination or horizontal transfer of APSE phage (Table S10, Supporting information). In spite of limited

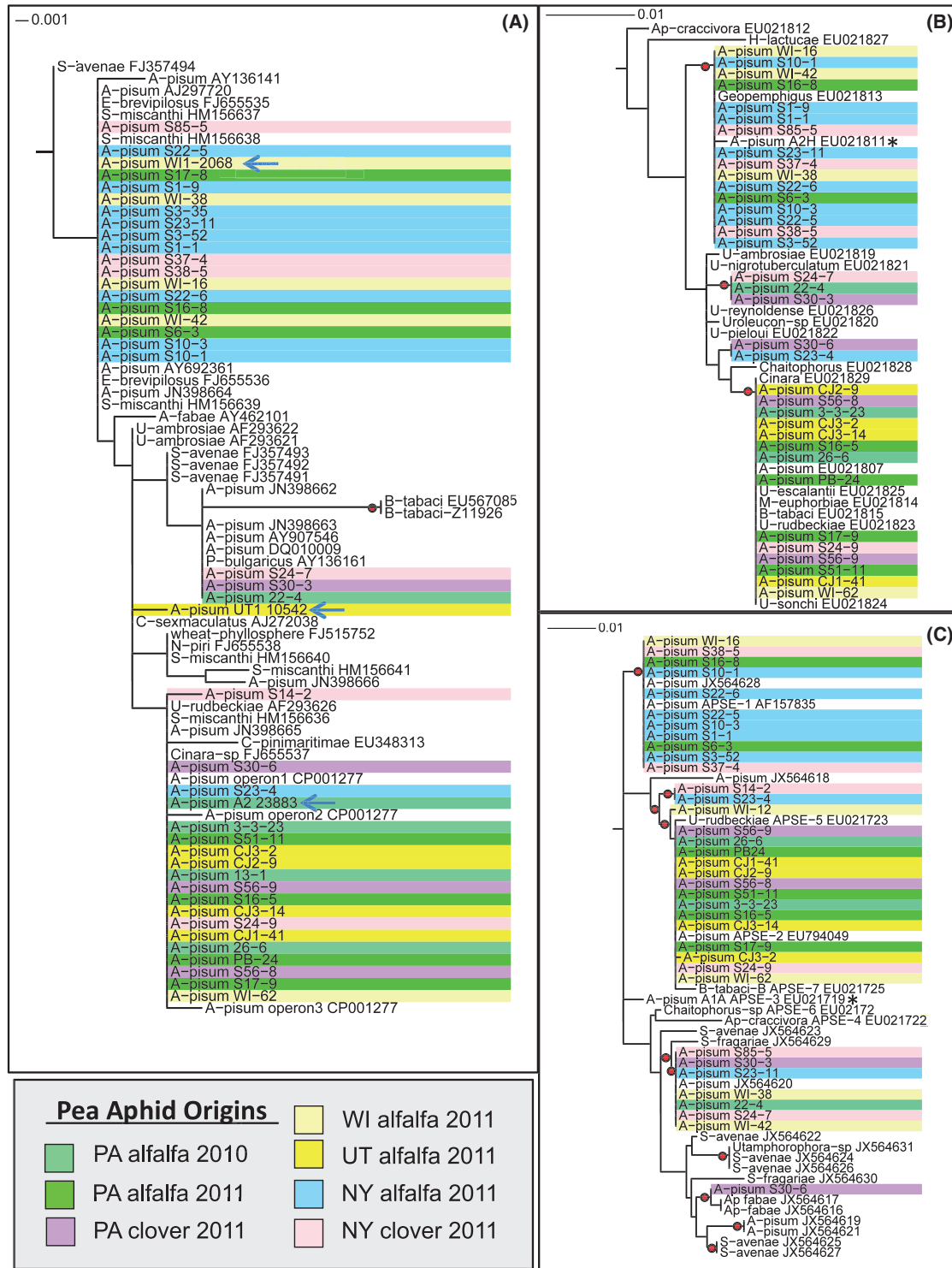


Fig. 2 Maximum likelihood phylogenies of *Hamiltonella defensa* and APSE phage-encoded genes. Shown here are phylogenies for portions of the (A) 16S rRNA and (B) *recJ* genes of *Hamiltonella defensa* and (C) the P3 gene of *H. defensa*-associated APSE phage. Alleles from different populations are indicated with different colour shading. Multiple allelic lineages within these populations illustrate the diversity of symbiont and phage strains. Note that all sequences here are named after their hosts and that '*' is used to illustrate the placement of genes from *H. defensa* and APSE of Utah aphids collected in 2003 (one representative sequence shown for three identical sequences generated from different clones of the same collection). Although based on small sample size, these symbionts are clearly distinct from those sampled in nearby Utah alfalfa fields in 2011. Arrows in panel 3a indicate the placement of representative sequences from 454 pyrosequencing. Finally, red circles are placed at nodes with bootstrap support values $\geq 80\%$.

sampling, we detected between two and four distinct genotypes in each of the six surveyed populations ($n = 3\text{--}9$ aphids sampled within populations).

Discussion

Diverse heritable symbiont assemblages found in North American A. pisum

Our survey of North American *A. pisum* populations using DGGE and 454 amplicon pyrosequencing revealed a diverse and dynamic assemblage of heritable bacteria infecting this aphid. Screening 318 total aphids from five populations identified nine heritable symbionts (including *Buchnera* and, for the first time, *Wolbachia*), as well as a number of additional lineages that are likely gut associates or contaminants from soil or plant material. While a substantial number of aphids were found to be free of secondary symbionts (ranging from 5.3 to 48% across populations), the majority of individual aphids in all populations were infected.

Sampled populations exhibited high symbiont diversity, with five to seven heritable bacterial species per population. *Hamiltonella defensa*, *R. insecticola* and *S. symbiotica* were found in all populations, often at high frequencies. The same was true for *Rickettsiella* symbionts, which were detected in N. American pea aphids for the first time. In contrast, *Spiroplasma*, X-type and *Rickettsia* were present at low levels (0–16%). When we consider prior findings of protective roles for each of these symbionts (Oliver *et al.* 2003; Scarborough *et al.* 2005; Guay *et al.* 2009; Łukasik *et al.* 2012), pea aphids potentially employ a diverse repertoire of microbially mediated defensive strategies in the field. Given that *Wolbachia* has not been previously reported in this well-studied aphid and that we detected it in only a single clone, it is unlikely that this heritable symbiont is a major player in the ecology and evolution of *A. pisum*.

Given that our 'universal' detection methods failed to identify other bacteria related to known heritable microbes, and since few of the other identified bacteria were prevalent, our findings suggest that no additional secondary symbionts with high infection frequencies and widespread distributions infect North American pea aphids. Nevertheless, for future studies of symbiont diversity, we suggest the use of universal 16S-based methods to determine the complement of common symbionts that might otherwise be missed with diagnostic PCR. Both DGGE and 454 amplicon pyrosequencing detected the same symbiont species, albeit with quantitative discrepancies between the two methods. Thus, the most practical approach for conducting large symbiont surveys may be to screen a representative sample with either of these universal methods to determine

the range of symbiont species present, before applying faster and more affordable diagnostic PCR screening to study the full collection.

Given that heritable symbionts are thought to be maintained in populations primarily through conferred benefits or reproductive manipulation (Bull 1983), the presence of diverse symbiont assemblages in parthenogenetic aphid populations indicates that infection likely has important consequences for aphid biology. These symbionts represent additional heritable genetic variation that may be acted on by natural selection. Indeed, rapid sweeps of symbionts throughout host species provide evidence for adaptive roles of heritable symbionts (in the field Jaenike *et al.* 2010; Himler *et al.* 2011). Field-based work is needed in the *A. pisum* system, however, to determine the importance of symbionts within natural populations, and the impact of environmental factors on symbiont-mediated phenotypes (Bensadia *et al.* 2006; Guay *et al.* 2009).

Infections with multiple secondary symbionts are common

In our survey, 25% of individual *A. pisum* harboured more than one secondary symbiont, and on average, individual aphids harboured 1.06 symbiont species aside from *Buchnera*. This number did vary among populations, however, ranging from 0.76 (PA vetch) to 1.74 (WI alfalfa). Infections with multiple secondary symbionts have been reported previously in pea aphids, including in a recent study, which detected 1.4 symbionts per aphid from European *A. pisum*, on average (Ferrari *et al.* 2012). Superinfections with (facultative) heritable symbionts are also common in other insects. In the whiteflies, *Bemisia tabaci* and *Trialeurodes vaporariorum*, for example, superinfection with combinations of *Arsenophonus*, *Cardinium*, *Hamiltonella*, *Rickettsia* and *Wolbachia* commonly occur (Chiel *et al.* 2007; Skaljic *et al.* 2010), and chestnut weevils (*Curculio sickimensis*) are reported to carry an impressive average of 3.18 symbionts per individual (Toju & Fukatsu 2011). Superinfection with multiple strains of the same symbiont species has also been reported (Kyei-Poku *et al.* 2005; Arthofer *et al.* 2009), with up to eight *Wolbachia* variants colonizing the same insect hosts (De-deine *et al.* 2005).

The prevalence of superinfections suggests that insects may conditionally benefit from particular combinations of symbionts, although the marginally significant rarity of superinfected pea aphids bearing *R. insecticola* suggests potential costs under some co-infection contexts, possibly due to host-harming competition between co-infecting microbes. Little is currently known, however, about costs and benefits of

superinfection in pea aphids. In one study, the fitness of pea aphids co-infected with *Serratia symbiotica* and *Rickettsia* appeared comparable to that of genetically identical aphids harbouring just one of these symbionts (Chen *et al.* 2000; Montllor *et al.* 2002). Another study found that aphids infected with *H. defensa* and *S. symbiotica* received more protection against parasitoids than when infected with either symbiont alone, yet superinfection resulted in substantial fitness costs (Oliver *et al.* 2006).

Given these fitness costs, it was surprising to note that *H. defensa* and *S. symbiotica* co-inhabited several aphid clones. In Utah, for example, 8/10 *S. symbiotica*-infected lines also harboured *H. defensa* ($n = 66$ surveyed aphids). These findings differed from those of a prior study (Oliver *et al.* 2006) reporting a rarity of *H. defensa* and *S. symbiotica* co-infections in Utah populations, plausibly due to detrimental effects of superinfection demonstrated in this same publication. These aphids were sampled in 2003 from alfalfa fields near to those used in our present study. Interestingly, the partial *recJ* and P3 sequences of *H. defensa* and APSE differed between UT aphids collected in 2003 vs. 2011 (Fig. 2b, c), albeit from a limited sample. It is, thus, possible that a shift in predominant symbiont strains or phage variants has influenced within-host interactions between these symbionts, permitting the spread of a previously rare superinfection. Other possibilities include strain evolution (vs. replacement) or the evolution of host-mediated traits affecting symbiont dynamics. Regardless, the observed shift in superinfection patterns hint that the costs and benefits of co-infection are dynamic within natural populations.

While we understand little about the causes and consequences of superinfection (see Appendix S1, Supporting information for more information), the various combinations of microbes within pea aphids greatly increase the diversity of cytoplasmic genotypes within this host. Should multiple infection commonly alter the phenotypes conferred by individual symbionts, it is possible that this variation is of functional and adaptive significance. Superinfection also creates opportunities for recombination and horizontal phage/gene transfer among co-infecting symbionts (Bordenstein & Wernegreen 2004), highlighting its likely importance for the population genetics and evolution of secondary symbionts.

Strain variation among secondary symbionts is an important source of diversity

While the number of heritable symbionts infecting natural *A. pisum* populations is impressive, little effort has been made to characterize the strain diversity of particular symbionts from this host. Through genotyp-

ing at a small number of genomic loci in our study, we detected strain variation for three secondary symbionts, revealing yet another source of symbiont-mediated cytoplasmic variation in this insect species.

Strain variation was most notable for *H. defensa*—when combining our results with those published previously, sequencing of the P3 gene (of APSE) revealed up to nine variants from pea aphids, compared to five and six variants, respectively, at *recJ* and 16S rRNA. The topologies of all three gene trees suggest that pea aphids have independently acquired *H. defensa* on several occasions. *Sitobion avenae* and *S. fragariae* also harboured multiple strains of *H. defensa*/APSE (and *R. insecticola*) based on inferences from the P3 and 16S rRNA phylogenies (Fig. 2). Thus, strain diversity may be a common feature of aphid secondary symbionts.

Other insect species have also been found to harbor multiple strains of some symbiont species. For instance, a recent study targeting three house-keeping genes from *Arsenophonus* of *B. tabaci* (whiteflies) found 19 haplotypes from six distinct phylogenetic lineages across 152 individuals from Europe and Africa (Mouton *et al.* 2012). Even greater variation was reported for *Wolbachia* across 20 populations of *Bryobia kissophila* (spider mites): genotyping at four loci identified 21 unique strains from just 35 individuals (Ros *et al.* 2012). While phylogenetics suggest that strain diversity can often arise due to independent, horizontal acquisition, recent findings of 14 monophyletic *Wolbachia* strains confined to the mosquito *Culex pipiens* suggest the potential for *in situ* diversification after invasion of hosts (Atyame *et al.* 2011).

Strain variation in heritable symbionts is likely important, as distinct strains can have different effects on host fitness and phenotypes. This has been seen, for instance, between two *Spiroplasma* strains infecting *Drosophila melanogaster*, which vary in their induction of the male-killing phenotype (Anbutsu & Fukatsu 2003). Co-infection with different CI-inducing *Wolbachia* strains can lead to bidirectional incompatibility due to different modification-rescue phenotypes; this phenomenon has been invoked in the process of *Nasonia* wasp speciation (Bordenstein *et al.* 2001). Phenotypic differences due to strain variation have also been reported for aphids. For example, *H. defensa* strains harbouring APSE-2 confer lower resistance to parasitic wasps compared to strains with APSE-3 (Oliver *et al.* 2005, 2009b), while strains of *R. insecticola* vary in their capacity to influence diet breadth (Ferrari *et al.* 2007) and resistance to parasitoids (Oliver *et al.* 2003; Vorburger & Gouskov 2011; Hansen *et al.* 2012). More generally, strains of protective symbionts could vary in the strength of protection conferred against particular genotypes of natural enemies (or even different species of enemy) (Łukasik *et al.* 2012; Schmid

et al. 2012). Given that natural enemies may possess the genetic variation required to overcome symbiont-mediated defense (Dion *et al.* 2011), this specificity could enable antagonistic coevolution between symbiont-protected hosts and natural enemies.

Variation in symbiont frequencies over space and time

In addition to variation created by the presence/absence of particular symbiont species, superinfection combinations and strain variation, we also found substantial variation in infection frequencies across pea aphid host races and geographic locations. While limited replication across food plant and geographical regions indicates a need for further exploration, our findings are consistent with previous studies (Montllor *et al.* 2002; Tsuchida *et al.* 2002; Leonardo & Muir 2003; Simon *et al.* 2003; Ferrari *et al.* 2004, 2012; Frantz *et al.* 2009). For instance, our trends of high *R. insecticola* frequencies in clover-feeding populations, and *H. defensa* enrichment in some alfalfa populations, have been previously reported (e.g. (Ferrari *et al.* 2004)).

Although our study identified suggestive instances of temporal frequency shifts, all were marginally significant before the Bonferroni correction, hinting at a need for increased sample sizes. Prior shifts in *S. symbiotica* frequency were found to correlate with increasing temperature (Montllor *et al.* 2002), meeting expectations from the laboratory given the known thermotolerance effect of this symbiont.

Overall, differences in symbiont frequency between host races, geographic locations or points in time could arise due to neutral processes (e.g. genetic drift; founder effects; acquisition of different symbionts since isolation) or due to the action of divergent selective pressures. Further work, with increased site replication within regions is needed to determine whether the geographic differences detected here persist across years or whether symbiont frequencies are more dynamic within regions. Similarly, temporal variation may also result from changes in selective pressures (e.g. natural enemies) throughout the season (Kwiatkowski & Vorburger 2012). But, again, replicated studies that also consider natural enemy prevalence and other environmental factors are needed to examine this.

Temporal and geographic variation in infection frequencies have, indeed, been reported for other heritable facultative symbionts. For instance, several bacteria from the whitefly *B. tabaci* differ in frequency in insects from different host plants and geographic locations, as well as between different host lineages (i.e. biotypes) (Pan *et al.* 2012). The chestnut weevil, *Curculio sikkimensis*, harbours six species of heritable symbiont, with notable frequency variation for several of these across

host plants and geographic locations (Toju & Fukatsu 2011). In this case, as seen for *R. insecticola* in pea aphids (Tsuchida *et al.* 2002), symbiont prevalence was found to correlate with climatic variables, namely temperature and precipitation. Geographic variation in *Arsenophonus* symbiont prevalence was also reported for the psyllid *Glycaspis brimblecombei*; interestingly, populations experiencing high rates of wasp parasitism showed higher levels of symbiont infection (Hansen *et al.* 2007). These latter studies suggest environmental factors that are likely to favour/disfavour symbiont infection in a natural setting and are consistent with non-neutral processes as important drivers of natural symbiont dynamics.

Conclusions

Our surveys of secondary symbionts across *A. pisum* have identified 28 unique cytoplasmic genotypes, based only on the combinations of heritable symbiont species found to infect single aphid individuals. When we consider the emerging picture of high strain diversity for some symbionts, it is clear that the cytoplasmically inherited genotypes of aphids are highly diverse (even when ignoring variation in mtDNA and *Buchnera* genomes). We propose that this diversity is maintained, in large part, due to balancing selection, with different symbiont species, strains and superinfections conferring advantages under a limited range of fluctuating environmental conditions. Furthermore, both horizontal (and sexual) transfer and maternal transmission failure are also of likely importance in the maintenance of this variation, so it is intriguing to note the growing number of avenues identified recently for the nonvertical transfer of secondary symbionts (Moran & Dunbar 2006; Caspi-Fluger *et al.* 2012; Gehrer & Vorburger 2012). More systematic and fine-scaled investigations of the dynamics of *A. pisum* heritable symbionts are needed, as are explorations of the selective and nonselective factors responsible for observed patterns. Given the prevalence of heritable symbionts across the arthropods, and the emerging trends of symbiont diversity and dynamism, there is clearly a broader need to quantify symbionts' contributions to their hosts' genetic diversity and the resulting implications for arthropod evolution.

Acknowledgements

The authors thank Anthony Ives, Ted Evans, Nick Tuttle and Rachae DiSciullo for assistance with aphid collections, and three anonymous reviewers for helpful suggestions on the manuscript. This work was funded by NSF grant #1050128 to KMO and #1050098 to JAR.

References

- Anbutsu H, Fukatsu T (2003) Population dynamics of male-killing and non-male-killing spiroplasmas in *Drosophila melanogaster*. *Applied and Environmental Microbiology*, **69**, 1428–1434.
- Arthofer W, Riegler M, Schneider D, Krammer M, Miller WJ, Stauffer C (2009) Hidden *Wolbachia* diversity in field populations of the European cherry fruit fly, *Rhagoletis cerasi* (Diptera, Tephritidae). *Molecular Ecology*, **18**, 3816–3830.
- Atyame CM, Delsuc F, Pasteur N, Weill M, Duron O (2011) Diversification of *Wolbachia* endosymbiont in the *Culex pipiens* mosquito. *Molecular Biology and Evolution*, **28**, 2761–2772.
- Bano N, Hollibaugh JT (2002) Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. *Applied and Environmental Microbiology*, **68**, 505–518.
- Bensadia F, Boudreault S, Guay JF, Michaud D, Cloutier C (2006) Aphid clonal resistance to a parasitoid fails under heat stress. *Journal of Insect Physiology*, **52**, 146–157.
- Bordenstein SR, Wernegreen JJ (2004) Bacteriophage flux in endosymbionts (*Wolbachia*): Infection frequency, lateral transfer, and recombination rates. *Molecular Biology and Evolution*, **21**, 1981–1991.
- Bordenstein SR, O'Hara FP, Werren JH (2001) *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature*, **409**, 707–710.
- Bull JJ (1983) *Evolution of Sex Determining Mechanisms*. Benjamin/Cummings, Menlo Park, California.
- Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.
- Caspi-Fluger A, Inbar M, Mozes-Daube N *et al.* (2012) Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated. *Proceedings of the Royal Society B-Biological Sciences*, **279**, 1791–1796.
- Chen DQ, Purcell AH (1997) Occurrence and transmission of facultative endosymbionts in aphids. *Current Microbiology*, **34**, 220–225.
- Chen DQ, Montllor CB, Purcell AH (2000) Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrtosiphon pisum*, and the blue alfalfa aphid, *A-kondoi*. *Entomologia Experimentalis Et Applicata*, **95**, 315–323.
- Chiel E, Gottlieb Y, Zchori-Fein E *et al.* (2007) Biotype-dependent secondary symbiont communities in sympatric populations of *Bemisia tabaci*. *Bulletin of Entomological Research*, **97**, 407–413.
- Cole JR, Chai B, Farris RJ *et al.* (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Research*, **33**, D294–D296.
- Darby AC, Douglas AE (2003) Elucidation of the transmission patterns of an insect-borne bacterium. *Applied and Environmental Microbiology*, **69**, 4403–4407.
- Darby AC, Tosh CR, Walters KFA, Douglas AE (2003) The significance of a facultative bacterium to natural populations of the pea aphid *Acyrtosiphon pisum*. *Ecological Entomology*, **28**, 145–150.
- Dedeine F, Ahrens M, Calcaterra L, Shoemaker DD (2005) Social parasitism in fire ants (*Solenopsis* spp.): a potential mechanism for interspecies transfer of *Wolbachia*. *Molecular Ecology*, **14**, 1543–1548.
- Degnan PH, Moran NA (2008) Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. *Molecular Ecology*, **17**, 916–929.
- Dion E, Zele F, Simon JC, Outreman Y (2011) Rapid evolution of parasitoids when faced with the symbiont-mediated resistance of their hosts. *Journal of Evolutionary Biology*, **24**, 741–750.
- Douglas AE (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology*, **43**, 17–37.
- Douglas AE (2009) The microbial dimension in insect nutritional ecology. *Functional Ecology*, **23**, 38–47.
- Duron O, Bouchon D, Boutin S *et al.* (2008) The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *Bmc Biology*, **6**, doi:10.1186/1741-7007-6-27
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
- 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**, 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**, 323–329.
- Ferrari J, West JA, Via S, Godfray HCJ (2012) Population genetic structure and secondary symbionts in host-associated populations of the pea aphid complex. *Evolution*, **66**, 375–390.
- Frantz A, Calcagno V, Mieuze L, Plantegenest M, Simon JC (2009) Complex trait differentiation between host-populations of the pea aphid *Acyrtosiphon pisum* (Harris): implications for the evolution of ecological specialisation. *Biological Journal of the Linnean Society*, **97**, 718–727.
- Gehrer L, Vorburger C (2012) Parasitoids as vectors of facultative bacterial endosymbionts in aphids. *Biology Letters*, **8**, 613–615.
- Genecodes (2003) *Sequencher*. Genecodes Co, Ann Arbor, Michigan.
- Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, **27**, 221–224.
- Guay JF, Boudreault S, Michaud D, Cloutier C (2009) Impact of environmental stress on aphid clonal resistance to parasitoids: role of *Hamiltonella defensa* bacterial symbiosis in association with a new facultative symbiont of the pea aphid. *Journal of Insect Physiology*, **55**, 919–926.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.
- Hansen AK, Jeong G, Paine TD, Stouthamer R (2007) Frequency of secondary symbiont infection in an invasive psyllid relates to parasitism pressure on a geographic scale in California. *Applied and Environmental Microbiology*, **73**, 7531–7535.
- Hansen AK, Vorburger C, Moran NA (2012) Genomic basis of endosymbiont-conferred protection against an insect parasitoid. *Genome Research*, **22**, 106–114.
- Haynes S, Darby AC, Daniell TJ *et al.* (2003) Diversity of bacteria associated with natural aphid populations. *Applied and Environmental Microbiology*, **69**, 7216–7223.
- 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 Microbiology Letters*, **281**, 215–220.
- Himler AG, Adachi-Hagimori T, Bergen JE *et al.* (2011) Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science*, **332**, 254–256.
- Ishak HD, Plowes R, Sen R *et al.* (2011) Bacterial diversity in *Solenopsis invicta* and *Solenopsis geminata* ant colonies characterized by 16S amplicon 454 pyrosequencing. *Microbial Ecology*, **61**, 821–831.
- Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ (2010) Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. *Science*, **329**, 212–215.
- Kwiatkowski M, Vorburger C (2012) Modeling the ecology of symbiont-mediated protection against parasites. *American Naturalist*, **179**, 595–605.
- Kyei-Poku GK, Colwell DD, Coghlin P, Benkel B, Floate KD (2005) On the ubiquity and phylogeny of *Wolbachia* in lice. *Molecular Ecology*, **14**, 285–294.
- Lee AH, Husseneder C, Hooper-Bui L (2008) Culture-independent identification of gut bacteria in fourth-instar red imported fire ant, *Solenopsis invicta* Buren, larvae. *Journal of Invertebrate Pathology*, **98**, 20–33.
- Leonardo TE (2004) Removal of a specialization-associated symbiont does not affect aphid fitness. *Ecology Letters*, **7**, 461–468.
- Leonardo TE, Muir GT (2003) Facultative symbionts are associated with host plant specialization in pea aphid populations. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, S209–S212.
- Łukasik P, van Asch M, Guo H, Ferrari J, Godfray HCJ (2012) Unrelated facultative endosymbionts protect aphids against a fungal pathogen. *Ecology Letters*, doi:10.1111/ele.12031.
- Miller M, Holder M, Vos R *et al.* (2012) http://www.phylo.org/sub_sections/porta.
- Montllor CB, Maxmen A, Purcell AH (2002) Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum* under heat stress. *Ecological Entomology*, **27**, 189–195.
- Moran NA, Baumann P (2000) Bacterial endosymbionts in animals. *Current Opinion in Microbiology*, **3**, 270–275.
- Moran NA, Dunbar HE (2006) Sexual acquisition of beneficial symbionts in aphids. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12803–12806.
- Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165–190.
- Mouton L, Thierry M, Henri H *et al.* (2012) Evidence of diversity and recombination in *Arsenophonus* symbionts of the *Bemisia tabaci* species complex. *Bmc Microbiology*, **12**, S10.
- Myers RM, Fischer SG, Lerman LS, Maniatis T (1985) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel-electrophoresis. *Nucleic Acids Research*, **13**, 3131–3145.
- 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 of the United States of America*, **100**, 1803–1807.
- Oliver KM, Moran NA, Hunter MS (2005) Variation in resistance to parasitism in aphids is due to symbionts not host genotype. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 12795–12800.
- Oliver KM, Moran NA, Hunter MS (2006) Costs and benefits of a superinfection of facultative symbionts in aphids. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 1273–1280.
- Oliver KM, Campos J, Moran NA, Hunter MS (2008) Population dynamics of defensive symbionts in aphids. *Proceedings of the Royal Society B-Biological Sciences*, **275**, 293–299.
- Oliver KM, Degnan PH, Burke GR, Moran NA (2009a) Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, **55**, 247–266.
- Oliver KM, Degnan PH, Hunter MS, Moran NA (2009b) Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science*, **325**, 992–994.
- 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**, 247–266.
- Pan HP, Li XC, Ge DQ *et al.* (2012) Factors affecting population dynamics of maternally transmitted endosymbionts in *Bemisia tabaci*. *PLoS ONE*, **7**, e30760.
- Ros VID, Fleming VM, Feil EJ, Breeuwer JAJ (2012) Diversity and recombination in *Wolbachia* and *Cardinium* from *Bryobia* spider mites. *Bmc Microbiology*, **12**, S13.
- 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-Biological Sciences*, **273**, 603–610.
- Russell JA, Latorre A, Sabater-Munoz B, Moya A, Moran NA (2003) Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Molecular Ecology*, **12**, 1061–1075.
- Sandström JP, Russell JA, White JP, Moran NA (2001) Independent origins and horizontal transfer of bacterial symbionts of aphids. *Molecular Ecology*, **10**, 217–228.
- Scarborough CL, Ferrari J, Godfray HCJ (2005) Aphid protected from pathogen by endosymbiont. *Science*, **310**, 1781.
- Schmid M, Sieber R, Zimmermann YS, Vorburger C (2012) Development, specificity and sublethal effects of symbiont-conferred resistance to parasitoids in aphids. *Functional Ecology*, **26**, 207–215.
- Simon JC, Carre S, Boutin M *et al.* (2003) Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proceedings of the Royal Society B-Biological Sciences*, **270**, 1703–1712.
- Simon JC, Boutin S, Tsuchida T *et al.* (2011) Facultative symbiont infections affect aphid reproduction. *PLoS ONE*, **6**, e21831.
- Skaljac M, Zanik K, Ban SG, Kontsedalov S, Ghanim M (2010) Co-infection and localization of secondary symbionts in two whitefly species. *BMC Microbiology*, **10**, 142.
- Toju H, Fukatsu T (2011) Diversity and infection prevalence of endosymbionts in natural populations of the chestnut weevil: relevance of local climate and host plants. *Molecular Ecology*, **20**, 853–868.
- Tsuchida T, Koga R, Shibao H, Matsumoto T, Fukatsu T (2002) Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, *Acyrtosiphon pisum*. *Molecular Ecology*, **11**, 2123–2135.

- Tsuchida T, Koga R, Fukatsu T (2004) Host plant specialization governed by facultative symbiont. *Science*, **303**, 1989.
- Tsuchida T, Koga R, Horikawa M *et al.* (2010) Symbiotic bacterium modifies aphid body color. *Science*, **330**, 1102–1104.
- Vorburger C, Gehrler L, Rodriguez P (2010) A strain of the bacterial symbiont *Regiella insecticola* protects aphids against parasitoids. *Biology Letters*, **6**, 109–111.
- Vorburger C, Gouskov A (2011) Only helpful when required: a longevity cost of harbouring defensive symbionts. *Journal of Evolutionary Biology*, **24**, 1611–1617.
- Werren JH, Baldo L, Clark ME (2008) Wolbachia: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, **6**, 741–751.
- Zwickl D (2006) *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion*. University of Texas, Austin, Texas.

J.R. and K.O. designed the research and wrote the manuscript. J.R. created figures, while K.O. and J.R. created tables. S.W., D.S., K.K., Y.H., P.L., S.D. and I.A. carried out all molecular work. M.N. and D.S. performed aphid collections.

Data accessibility

Sanger sequences are deposited in GenBank under the accession numbers KC206059–KC206062, KC215465–KC215470 and KC242466–KC242612. Pyrosequencing reads are deposited at the GenBank Short Read Archive under accession number SRA062093.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Example Denaturing Gradient Gel Electrophoresis (DGGE) result.

Fig. S2 Proportions of 454 16S rRNA sequence reads corresponding to various bacteria.

Fig. S3 16S rRNA haplotypes for three secondary symbionts at likely polymorphic sites.

Fig. S4 Maximum likelihood phylogeny of 16S rRNA genes of *Regiella*.

Fig. S5 Maximum likelihood phylogeny of 16S rRNA genes of *Rickettsia*.

Table S1 Representative 16S rRNA sequences from DGGE reactions.

Table S2 PCR primers and reaction conditions used in this study.

Table S3 Summary of 454 sequencing results.

Table S4 Results of secondary symbiont surveys across the five sampled populations.

Table S5 Statistical analyses comparing symbiont frequencies across populations.

Table S6 Statistical analyses comparing symbiont frequencies over time in the PA alfalfa population.

Table S7 Statistical analyses on the numbers of symbionts per aphid across populations.

Table S8 Observed and expected numbers of aphids harboring specific single, double, triple, and quadruple infections.

Table S9 Observed and expected numbers of aphids harboring particular pairwise combinations of symbionts.

Table S10 Genotypes of *Hamiltonella defensa* and APSE from NY, PA, WI, and UT aphids.

Appendix S1 Additional discussion and details on experimental methodologies can be found in this section.