Bulletin of Entomological Research (2015) **105**, 110–120
doi:10.1017/S0007485314000820

Armored scale insect endosymbiont diversity at the species level: genealogical patterns of *Uzinura diasipipdicola* in the *Chionaspis pinifoliae–Chionaspis heterophyllae* species complex (Hemiptera: Coccoidea: Diaspididae)

J.C. Andersen[1], R.A. Gwiazdowski[2], K. Gdanetz[3] and M.E. Gruwell[4]*

[1]Department of Environmental Science Policy and Management, University of California, Berkeley, CA 94720, USA: [2]Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada: [3]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA: [4]Penn State Erie, The Behrend College, School of Science, Erie, PA 16563, USA

Abstract

Armored scale insects and their primary bacterial endosymbionts show nearly identical patterns of co-diversification when viewed at the family level, though the persistence of these patterns at the species level has not been explored in this group. Therefore we investigated genealogical patterns of co-diversification near the species level between the primary endosymbiont *Uzinura diaspidicola* and its hosts in the *Chionaspis pinifoliae–Chionaspis heterophyllae* species complex. To do this we generated DNA sequence data from three endosymbiont loci (rspB, GroEL, and 16S) and analyzed each locus independently using statistical parsimony network analyses and as a concatenated dataset using Bayesian phylogenetic reconstructions. We found that for two endosymbiont loci, 16S and GroEL, sequences from *U. diaspidicola* were broadly associated with host species designations, while for rspB this pattern was less clear as *C. heterophyllae* (species S1) shared haplotypes with several other *Chionaspis* species. We then compared the topological congruence of the phylogenetic reconstructions generated from a concatenated dataset of endosymbiont loci (including all three loci, above) to that from a concatenated dataset of armored scale hosts, using published data from two nuclear loci (28S and EF1α) and one mitochondrial locus (COI–COII) from the armored scale hosts. We calculated whether the two topologies were congruent using the Shimodaira–Hasegawa test. We found no significant differences ($P = 0.4892$) between the topologies suggesting that, at least at this level of resolution, co-diversification of *U. diaspidicola* with its armored scale hosts also occurs near the species level. This is the first such study of co-speciation at the species level between *U. diaspidicola* and a group of armored scale insects.

*Author for correspondence
Phone: 814-898-6276
Fax: 814-898-6213
E-mail: meg26@psu.edu

Keywords: genealogical concordance, endosymbiosis, pine-feeding, North America, TreeMap

(Accepted 29 September 2014; First published online 26 November 2014)

Introduction

Many groups of Sternorrhynchan insects depend on endosymbiotic bacteria to provide nutrients not present in their plant-fluid based diets (Buchner, 1965; Moran, 2001; von Dohlen *et al*., 2001; Baumann, 2005; Moran *et al*., 2008). Due in part to the obligate associations between these organisms, phylogenetic analyses of these primary-endosymbionts and their insect hosts have found patterns of phylogenetic congruence at the family level (Gruwell *et al*., 2007; Urban & Cryan, 2012; Kuechler *et al*., 2013) as well as at the species level (Ahmed *et al*., 2013; Liu *et al*., 2013). One group of Sternorhynchans where family level patterns of phylogenetic congruence between host and primary endosymbiont have been observed is the armored scale insects (Coccoidea: Diaspididae) and their primary endosymbiont *Uzinura diaspidicola* Gruwell *et al*. (2007). However, patterns of phylogenetic congruence between these insects and *U. diaspidicola* remain to be explored below the family level.

When Gruwell *et al*. (2007) identified the primary endosymbiont of all armored scales as *U. diaspidicola* by comparing the family level phylogeneies of *U. diaspidicola* and its armored scale hosts, the authors included eleven genera of armored scales that were each represented by two or more species. Seven of these genera showed strict co-diversification with their endosymbionts, whereas patterns for the remaining genera were less congruent. All genera, however, were represented by only a small number of individuals per genus, and thus it remains unclear whether or not the previously observed family level congruencies also extend to the species level for armored scales. One of the multi-species genera included by Gruwell *et al*. (2007), *Chionaspis*, and two pine-feedings members of this genus *C. pinifoliae* Fitch and *C. heterophyllae* Cooley were recently the subject of an extensive intra-specific study (Gwiazdowski *et al*., 2011). In that study, Gwiazdowski *et al*. (2011) sampled 320 individuals from across their North American range and found that North American species of *Chionaspis* formed a monophyletic group, confirmed *C. heterophyllae* as a single species, and revealed *C. pinifoliae* to be a complex of at least nine species-based on a morphological survey in concert with multi-locus genealogical concordance. Four species from this group were recently described (Vea *et al*., 2012), and the extensive geographic sampling for this group provides a fine-scale context to explore host/endosymbiont co-evolution near the species level in armored scales.

Therefore, we took advantage of this existing dataset and we built upon the host DNA sequence data collected by Gwiazdowski *et al*. (2011) by sequencing three loci from the primary endosymbiont *U. diaspidicola* to examine whether patterns exist of phylogenetic congruence between closely related species in the *C. pinifoliae–C. heterophyllae* species complex and their primary endosymbiont. We then compared these phylogenies to the biogeographic and plant–host information associated with each individual to explore whether patterns of diversity of *U. diaspidicola* and/or its hosts in the *C. pinifoliae–C. heterophyllae* species complex are shaped by biogeographic or plant–host association processes.

Materials and methods

Sampling and gene selection

Full collection information including locality, species designations, and per-specimen GenBank accession numbers for all host DNA sequences for all specimens used in this study are available in Supplementary Table 1 of Gwiazdowski *et al*. (2011). Though Vea *et al*. (2012) have recently described several species in the *C. pinifoliae–C. heterophyllae* species complex, we follow the species designation scheme of Gwiazdowski *et al*. (2011; e.g., S1 for *C. heterophyllae*, S2 for *Chionaspis* species number two, S3 for *Chionaspis* species number three, etc.) for easier comparison between these two studies.

DNA extraction and PCR amplification

Total genomic DNA from individual armored scale insects was isolated as part of Gwiazdowski *et al*. (2011) using the DNEasy Blood and Tissue Kit (Qiagen, Valencia, California). These extracts were used to amplify fragments of DNA from *U. diaspidicola* following PCR protocols for armored scale insect endosymbionts (Gruwell *et al*., 2007, 2009). We amplified three endosymbiont genes from *U. diaspidicola*: 16S rRNA, GroEL (a chaperone gene), and rspB (a nuclear protein-coding gene); the amino acid variation in these genes is consistent with active genes mapped in the *U. diaspidicola* genome (Sabree *et al*., 2013).

Individual gene fragments were amplified using standard PCR procedures on a TC – 3000 G thermal cycler (Techne Corp, MN) using the following protocols. Previously published primers were used to amplify fragments of 16S, and a combination of previously published primers and novel primers developed for this study were used to amplify fragments of rspB and GroEL (table 1). Reactions were performed using Nexus PCR Premix Taq following the manufacturer's protocol (Bionexus, Oakland, California) and brought to 25 µl using ultra pure H_2O. Thermocycler conditions were as follows: for 16S and GroEL genes, an initial denaturation temperature of 95 °C for 5 min, followed by 30 cycles of 95 °C for 30 s, 49 °C for 60 s, and 72 °C for 30 s, and finishing with a final extension at 72 °C for 5 min; and for rspB, an initial denaturation temperature of 95 °C for 5 min, followed by 5 cycles of 95 °C for 30 s, 54 °C for 30 s, and 72 °C for 1 min, 5 cycles of 95 °C for 30 s, 51 °C for 30 s, and 72 °C for 1 min, 5 cycles of 95 °C for 30 s, 49 °C for 30 s, and 72 °C for 1 min, 5 cycles of 95 °C for 30 s, 47 °C for 30 s, and 72 °C for 1 min, 30 cycles of 95 °C for 30 s, 45 °C for 30 s, and 72 °C for 1 min, finishing with a final extension at 72 °C for 10 min (Moulton & Wiegmann, 2004). PCR products were visualized using 1.5% agarose gels stained with EZVision® (Amresco, Solon, Ohio).

DNA sequencing and alignment

PCR products were either purified with ExoSAP-IT® (Affymetrix, Santa Clara, California) and sequenced at Penn State Nucleic Acid Facility (University Park, Pennsylvania),

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press

Table 1. Names, orientations, and sequences for primers used in this study.

Locus	F/R[1]	Name	Sequence
16S	F	16SA1[2]	ATA GTT TGA TCM TGG CTC AG
	F	s30BUCH16S[2]	GGC GGC AAG CCT AAC ACA TGC AAG T
	F	s688DIASP[2]	GGA ATG TAT GGT GTA GCG GTG AAA T
	R	8Fbac[3]	AGA GTT TGA TCC TGG CTC AG
	R	A1271DIASP[2]	CAT TGT AGC ACG TGG GTA GCC CAA G
	R	A1446 BUCH 16S[2]	CTC CCA TGG TGT GAC GGG CGG TGT G
	R	16SB1[2]	TAC GGY TAC CTT GTT ACG ACT T
	R	1495R[4]	CTA CGG CTA CCT TGT TAC GA
rspB	F	UZIrspBF1	TTG CAG GGA TAG GAT AAT CTA C
	F	rspB55F[5]	GAA GAA TCA TCA TTT GCA GGT ATA GG
	R	UZIrspBR1	AAT TTC AAT TTT CTA AAC G
	R	rspB60+R[5]	CTT ACT AGA GAT CCA AGA ATG GTA GA
GroEL	F	UZIGroELF1	GGT CTT AAA AAT GTA GCT GCT GG
	F	GroEL2225FUZ[5]	CTA ATG ACG TAG CTG GAG AT G
	R	UZIGroELR1	CAG AAG TAG TAT TAT CTA TTT GAG
	R	GroEL1050RUZ[5]	CAA TAC TGC CAC TCC ACC TGC

[1]Designation of forward (F) or reverse (R) primers.
[2]Primers originally published in Gruwell *et al.* (2007).
[3]Universal primer originally published in Gruwell *et al.* (2010).
[4]Endosymbiont specific primer derived from Weisburg *et al.* (1991).
[5]Primers originally published in Andersen *et al.* (2014).

or purified and sequenced at the High throughput Genomics Center (Seattle, Washington). Sequences were edited and assembled using Geneious v. 5.6.2 (Drummond *et al.*, 2012), and sequence alignments were created using MUSCLE (Edgar, 2004). All alignments were then truncated to the length of the shortest assembled sequence. DNA sequences generated in this study are available on Genbank under the following accession numbers: for 16S KF300549–KF300621, for GroEL KF300622–KF300699, and for rspB KF300700–KF300768. A concatenated dataset was then constructed from the endosymbiont sequence data and trimmed to only include individuals from whom at least two of the three-endosymbiont loci were available. Insect DNA sequences from the nuclear loci 28S and elongation factor 1-alpha (EF1α) and the mitochondrial locus cytochrome oxidase I and II (COI–COII) for all individuals from which at least two endosymbiont sequences were sequenced were then downloaded from GenBank, and each locus was aligned independently using MUSCLE (Edgar, 2004). The alignment for each locus was then truncated and a concatenated dataset was generated as per the endosymbiont dataset. No regions of the endosymbiont datasets were masked from analyses, as there were no insertions or deletions present. For the concatenated host dataset there were no insertions or deletions present in 28S; however, the intergenic region located between COI and COII was excluded from analysis, as were the introns in EF1α.

Genetic analyses

Statistical parsimony networks were constructed for each of the endosymbiont alignments using TCS v. 1.21 (Clement *et al.*, 2000) with a 95% connection limit. We then performed phylogenetic reconstructions for both the concatenated endosymbiont dataset and the concatenated host dataset using MrBayes (Huelsenbeck, 2001) with the following steps. First, the concatenated datasets were partitioned by gene and further partitioned by codon for each of the protein coding genes (GroEL, rspB, EF1α and COI–COII). Then a best-fitting nucleotide model for each gene was determined based on AIC scores using ModelTest (Posada & Crandall, 1998), and the best-fit model for each gene was then used to assign the appropriate number of unique substitution rates and model of rate variation to each partition. For both datasets two independent runs, each with four chains, were analyzed for 20 million generations with sampling every 1000th generation and a heating of 0.2. All phylogenetic analyses were run through the CIPRES Science Gateway (Miller *et al.*, 2010). The program Tracer (Rambaut & Drummond, 2007) was used to visualize that the log-likelihood scores and a burn-in of 20% (4 million generations) was performed before summarizing the majority rule consensus tree for each dataset.

To compare the congruence of the majority rule consensus trees constructed from the concatenated endosymbiont and the concatenated host datasets we compared the fit of the two trees to the endosymbiont dataset and tested whether there were significant differences between the trees using a Shimodaira-Hasegawa (SH) test as implemented in the phylogenetic program PAUP*4.0beta (Swofford, 2002). Significance was obtained using 10,000 bootstrap replicates.

Biogeographic and plant–host association analyses

Collection information, including the geographic region of origin and the plant species from which individual specimens were collected from were scored as characters in MacClade v. 4.08 (Maddison & Maddison, 2005). For the biogeographic analyses individuals collected in Georgia, Massachusetts, Maine, North Carolina, Tennessee and Wisconsin were scored as 'Eastern US', individuals collected in Arizona, Baja California, California, Colorado, Texas, Utah and Washington State were scored as 'Western US and Baja CA,' and individuals collected in the Mexican states of Oaxaca, Sonora and Mexico were scored as 'Mexico.' For the plant association analyses, plant host species were scored by their subgenus affiliations following the Gymnosperm Database (Earle, 2014) and the Jepson Herbarium eFlora Database (Jepson Flora Project 2014). These characters were then mapped onto the host and endosymbiont phylogenies using the trace character feature in MacClade, and the ancestral state for each node was calculated using maximum parsimony reconstruction.

Results

The best-fit model for each locus are as follows: for 16S the best-fit model was TIM+I, for GroEL the best-fit model was F81

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press

Table 2. Alignment length and character summary status for each locus analyzed. The overall length of the truncated alignment for both endosymbiont and host loci, the number of constant sites (CS) and the number of parsimony-informative sites (PI) as calculated in PAUP (Swofford, 2002) are reported.

Genome	Locus	Length (bp)	CS	PI
Uzinura diaspidicola	16S	898	889	5
Uzinura diaspidicola	GroEL	594	573	12
Uzinura diaspidicola	rspB	514	500	10
Host – mtDNA	COI–COII	714	492	146
Host – nuDNA	EF1α	594	564	23
Host – nuDNA	28S	483	474	5

Table 3. Summary of collection information presented in Gwiazdowski *et al.* (2011) including 16 s, GroEL and rspB haplotype information. For those species that have been described, species designations (S1–S10) are followed by their scientific name.[1]

Location	Country	State	Host species	16S Haplotype	GroEL Haplotype	rspB Haplotype
Chionaspis gleditsiae						
D0932	US	MA	*Gleditsia triacanthos*			D0932B
S1 – *Chionaspis heterophyllae*						
D0941	US	CT	*Pinus nigra*	D1667A		
D0970	US	FL	*Pinus palustris*		D0970A	
D0977	US	FL	*Pinus elliottii*		D0970A	
D1081	US	NC	*Pinus taeda*	D1667A	D0970A	
D1085	US	NC	*Pinus taeda*			D1085B
D1505	US	VA	*Pinus taeda*	D1667A	D0970A	
D1506	US	VA	*Pinus taeda*		D1506A	
D1507	US	NC	*Pinus taeda*	D1667A	D0970A	D1085B
D1510	US	GA	*Pinus taeda*	D1667A	D0970A	
D1667	US	NC	*Pinus pungens*	D1667A		
D1670	US	GA	*Pinus taeda*	D1670B		
D1687	US	NY	*Pinus sylvestris*		D0970A	
D1820	US	FL	*Pinus elliottii*			D1565A
S2						
D1711	MX	OAX	*Pinus teocote*		D1711B	D1711B
D1721	MX	MEX	*Pinus devoniana*	D1721A		D1721A
D1726	MX	PUE	*Pinus pseudostrobus*			D1726B
D1735	MX	QUE	*Pinus cembroides*			D1085B
D1755	MX	JAL	*Pinus douglasiana*			D1085B
D1772	MX	DUR	*Pinus lumholtzii*			D1085B
S2 – *Chionaspis sonorae*						
D1780	MX	SON	*Pinus engelmannii*	D1780A	D1780A	D1085B
S3						
D1699	MX	CHP	*Pinus undet*		D1699A	
D1700	MX	CHP	*Pinus undet*	D1700B		
S4 – *Chionaspis caudata*						
D1703	MX	OAX	*Pinus patula*	D1703A	D1703A	
S5						
D1705	MX	OAX	*Pinus pseudostrobus*			D1705A
S6						
D1563	US	CA	*Pinus strobiformis*		D1506A	D1563A
D1579	US	CA	*Pinus lambertiana*			D1579B
D1601	US	CA	*Pinus ponderosa*			D1579B
D1602	US	CA	*Pinus jeffreyi*		D1506A	D1602A
D1605	US	CA	*Pinus ponderosa*			D1579B
D1784	MX	BCN	*Pinus quadrifolia*		D1506A	
D1789	MX	BCN	*Pinus contorta*		D1506A	D1579B
D1791	MX	BCN	*Pinus quadrifolia*	D0465A	D1506A	D1602A
D2200	MX	BCN	*Pinus attenuata*		D1506A	
D2207	US	CA	*Pinus jeffreyi*		D1506A	D1579B
D2208	US	CA	*Pinus jeffreyi*		D1506A	
D2211	US	CA	*Pinus sabiniana*		D1506A	
S6 – *Chionaspis torreyanae*						
D2235	US	CA	*Pinus torreyana*		D1506A	D1579B
S7						
D0147	US	MA	*Pinus strobus*		D0147A	D0147A
D0931	US	NJ	*Pinus strobus*	D0465A		

Continued

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press

Table 3. (*Cont.*)

Location	Country	State	Host species	16S Haplotype	GroEL Haplotype	rspB Haplotype
D1631	US	MA	*Pinus mugo*		D0147A	D1565A
D1633	US	MA	*Pinus strobus*		D0147A	D0147A
D1634	US	MA	*Pinus strobus*		D0147A	D0147A
D1637	US	MA	*Pinus strobus*			D0147A
D1638	US	MA	*Pinus strobus*			D0147A
D1643	US	MA	*Pinus strobus*	D0465A	D0147A	D0147A
D1649	US	WI	*Pinus strobus*	D0465A	D0147A	D0147A
D1650	US	TN	*Pinus strobus*		D0147A	D0147A
D1664	US	VA	*Pinus strobus*		D0147A	
S8						
D0465	US	CA	*Pinus contorta*	D0465A		D0465A
D1573	US	CA	*Pinus radiata*		D1573A	D0465A
D1585	US	CA	*Pinus contorta*	D0465A	D1573A	
D1594	US	CA	*Pinus contorta*		D1573A	
D2241	US	CA	*Pinus attenuata*	D0465A	D1573A	
D2242	US	CA	*Pinus ponderosa*		D1573A	
S9						
D1541	US	AZ	*Pinus discolor*	D1541A		
S10						
D0466	US	CA	*Pinus lambertiana*		D0466A	
D1148	US	CA	*Pinus attenuata*	D0465A	D0466A	
D1404	US	WA	*Pinus ponderosa*	D0465A	D0466A	
D1415	US	WA	*Pinus ponderosa*	D0465A	D0466A	
D1417	US	WA	*Pseudotsuga menziesii*	D0465A		
D1418	US	WA	*Pinus contorta*		D0466A	
D1516	US	TX	*Pinus discolor*			D1516A
D1517	US	TX	*Pinus eudlis*	D0465A	D0466A	D1517A
D1520	US	AZ	*Pinus leiophylla*		D0466A	
D1526	US	AZ	*Pinus ponderosa*		D0466A	
D1528	US	AZ	*Pinus strobiformis*	D0465A	D0466A	
D1535	US	AZ	*Pinus ponderosa*		D0466A	
D1536	US	AZ	*Pinus strobiformis*		D0466A	
D1547	US	AZ	*Pinus leiophylla*		D0466A	
D1551	US	AZ	*Pinus ponderosa*	D0465A		D1517A
D1565	US	CA	*Pinus monophylla*			D1565A
D1566	US	CA	*Pinus monophylla*	D0465A	D0466A	D1565A
D1567	US	CA	*Pinus monophylla*	D0465A		
D1570	US	CA	*Pinus lambertiana*	D0465A	D0466A	D1565A
D1577	US	OR	*Pinus jeffreyi*			D1517A
D1615	US	UT	*Pinus monophylla*	D0465A	D1615A	
D1621	US	CO	*Pinus ponderosa*	D0465A		D1565A
D1628	CA	ON	*Pinus resinosa*	D0465A		
D1654	US	ME	*Pinus mugo*		D0147A	D1565A
D1661	US	VT	*Pinus sylvestris*	D0465A		
D1734	MX	QUE	*Pinus pinceana*			D1734A
D1790	MX	BCN	*Pinus jeffreyi*			D1517A
D2205	US	CA	*Pinus monophylla*	D0465A	D1615A	D0465A
D2214	US	CA	*Pinus lambertiana*	D0465A		

[1]Vea *et al.* (2012) recently described three species included in this analysis, *C. caudata*, *C. sonorae*, and *C. torreyanae*. *Chionaspis sonorae* includes some individuals previously described as S2 and *C. torreyanae* includes some individuals previously described as S6 by Gwiazdowski *et al.* (2011). During analyses we have used S2 and S6 as per Gwiazdowski *et al.* (2011); however, in the table above we show both designations for reference.

+I, for rspB the best-fit model was TrN+I, for 28S the best-fit model was K81uf+I, for EF1α the best-fit model was TrNef+I, and for COI–COII the best-fit model was TrN+I+G. The alignment length for each locus, the number of constant sites, and the number of parsimony informative (PI) sites were calculated using PAUP*4b10.0 (Swofford, 2002), and are presented in table 2. Haplotype information for each locus, and summary collection information for all individuals are shown in table 3. The network analysis of the 16S rRNA data set included 36 sequences (fig. 1), the network analysis of the rspB data set included 43 sequences (fig. 1), and the network analysis of the GroEL data set included 52 sequences (fig. 1). These analyses showed broad patterns of congruence between haplotype groups and species in the *C. pinifoliae–C. heterophyllae* species complex as classified by Gwiazdowski *et al.* (2011) as there were no species with polyphyletic haplotypes for 16S, and only two species with polyphyletic haplotypes for GroEL and rspB (fig. 1). For 16S, the network diagram was divided into four sections each separated by two or more base pair changes. The first included all individuals of S1, the second included all individuals of S3, and S5, the third included all individuals of S2, and the final section included all individuals of species S4,

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press

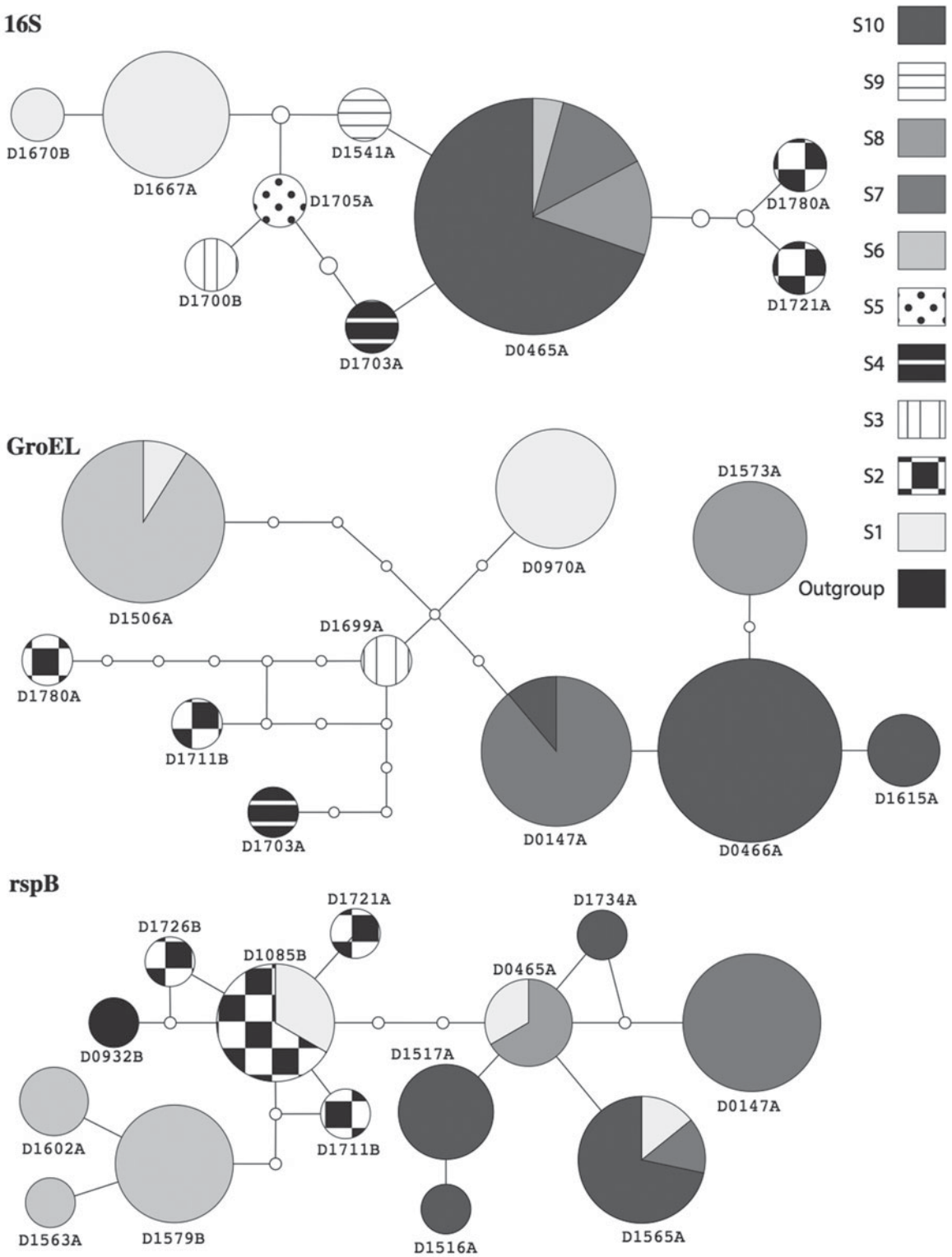

Fig. 1. Network diagram for 16S, GroEL, and rspB showing relationships between haplotypes. Haplotypes are drawn proportional to the number of individual sequences belonging to each group. The haplotype names adjacent to each circle represents the haplotype designations seen in table 3. Patterns are used to represent the species designation for the individual from whom each sequence was obtained. Distances between haplotypes are shown with lines and open circles (i.e., a line connecting two haplotypes equal one base-pair change, while open circles represent additional base pair changes).

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press

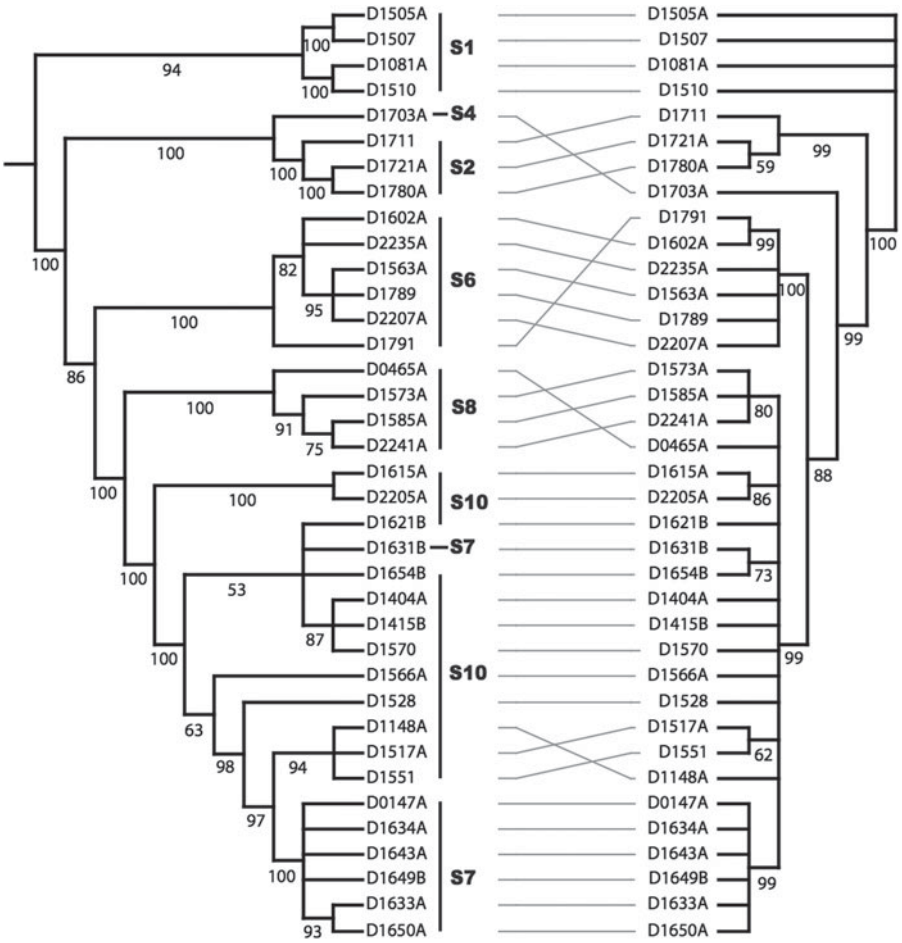

Fig. 2. Tanglegram comparing the majority rule consensus trees from the Bayesian reconstructions of the concatenated host (left) and concatenated endosymbiont (right) datasets. Numbers below each branch correspond with the BPP for each supported branch. To the right of the taxon labels for the host phylogeny are the species designations from Gwiazdowski *et al.* (2011). Lines connecting the taxon labels between the two phylogenies have been drawn to emphasize topological differences between the two phylogenies.

S6, S7, S8, S9 and S10. For GroEL, the network was divided into four sections separated by at least three basepair changes. The first grouping included all individuals of S7, S8 and S10. The second group included a single haplotype that included all the individuals of S6 as well as one individual of S1. The third grouping included a single haplotype that included only individuals of S1. The fourth group included four distantly related haplotypes each represented by a single individual only, two from S2, one from each S3 and S4. For rspB, the network was divided into three sections separated by at least three base pair changes. The first included individuals of Species S7, S8 and S10 as well as several individuals from S1. The second grouping included all individuals of S2, the one outgroup individual and several individuals of S1. The final grouping included only individuals of S6.

The Bayesian reconstruction of the concatenated endosymbiont and host datasets included sequences from 37 individuals (fig. 2). This majority rule consensus tree for the analysis of the concatenated endosymbiont dataset included monophyletic assemblages for S1, S2 and S6, with species S4 being putatively monophyletic (it was only represented by a single individual), and with species S6, S7, S8 and S10 forming a single unresolved clade. The majority rule consensus tree for the analysis of the host dataset reconstructed monophyletic species assemblages for S1, S2, S6 and S8 with species S4 being putatively monophyletic (again only represented by one individual), and with species S7 being reconstructed as polyphyletic and species S10 as paraphyletic, though together they form a well-supported clade with 100% Bayesian Posterior Probability (BPP).

The primary differences between the endosymbiont and host phylogenies were that the level of resolution seen in the host dataset was much greater than seen in the endosymbiont dataset, and that the placement of species S4 differed between reconstructions (fig. 2). In the endosymbiont phylogeny this species formed a clade including species S6–S10, while in the host phylogeny this species formed a clade only including species S2. These relationships were strongly supported in both datasets (endosymbiont – 99% BPP, host – 100% BPP). When we compared the congruence of the two trees using the SH test, we found no significant differences (ln difference $= -1.16733$, $P = 0.4892$) between the topologies of the COI–COII tree and the endosymbiont tree.

The biogeographic analyses for the host and endosymbiont datasets showed clear associations between the phylogenetic placements of individuals and geographic regions of collection (fig. 3). Two individuals (D1631B and D1654B) were collected

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press

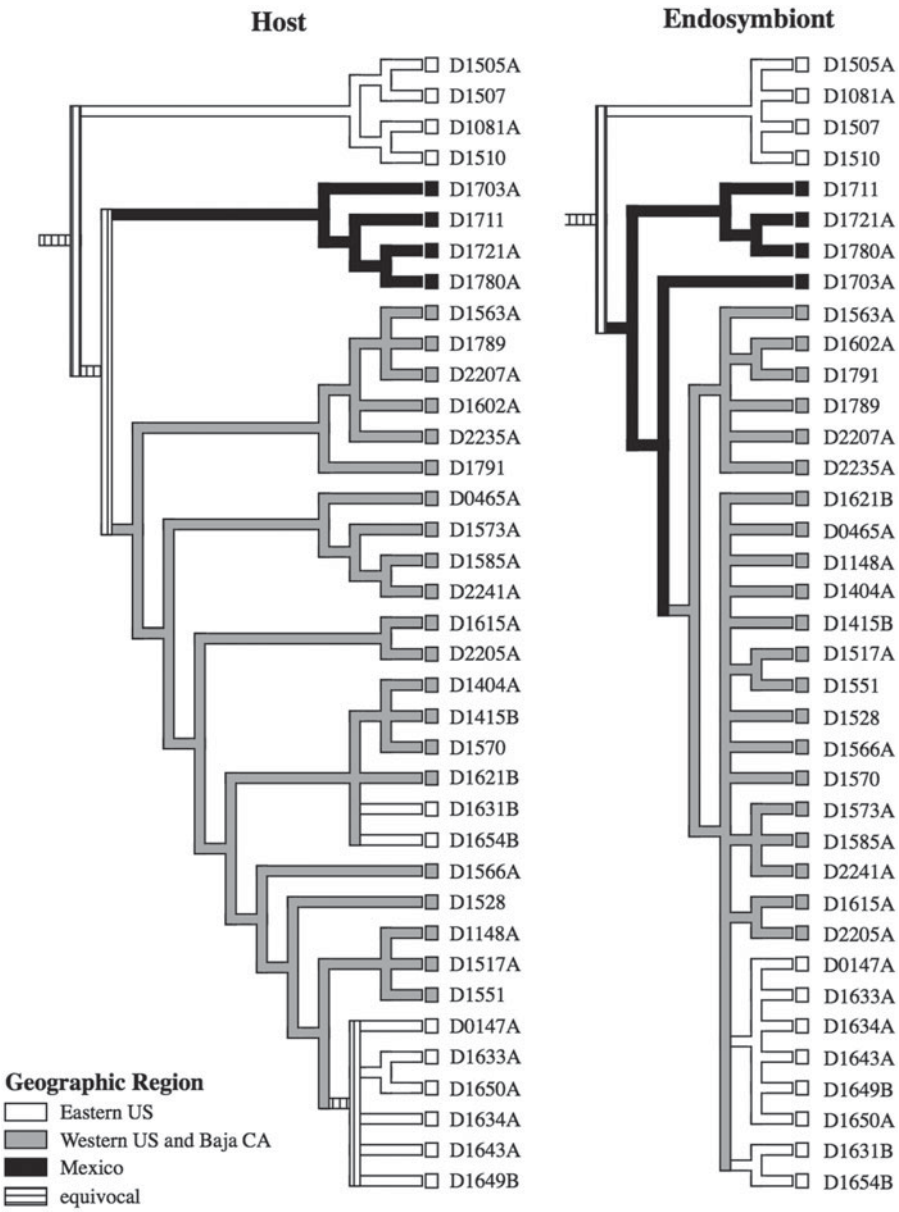

Fig. 3. Biogeographic patterns based on collection information published in Gwiazdowski *et al.* (2011). Geographic regions for the collection locality were mapped onto the host (left) and endosymbiont (right) phylogenies using the trace character feature in MacClade v 4.08 (Maddison & Maddison, 2005).

in the eastern US but group with individuals collected in the western US and Baja California. These two individuals were collected from an ornamental plant, *Pinus mugo*, and their collection in the eastern US could be the result of transportation of nursery stocks. The results from the plant association analysis found no clear patterns of phylogenetic associations with *Pinus* subgenera, though for both datasets we reconstructed the ancestral hosts of members of the *C. pinifoliae–C. heterophyllae* species complex to belong to the *Pinus* subgenus *Pinus* (fig. 4).

Discussion

Here we have explored patterns of co-diversification between members of the *C. pinifoliae–C. heterophyllae* species complex and their primary endosymbiont *U. diaspidicola*. We find no significant differences between the two phylogenies reconstructed for the host and endosymbiont datasets using a SH test ($P = 0.4892$). This congruence suggests that insect and endosymbiont co-evolution seen at the family level in armored scale insects (Gruwell *et al.*, 2007) may also occur at the species level among members of the *C. pinifoliae–C. heterophyllae* species complex. Similar species level patterns have recently been observed in other Sternorrhynchan-endosymbiont systems where phylogenetic patterns of co-diversification between closely related insect species and their endosymbionts have been observed in white flies (Ahmed *et al.*, 2013) and aphids (Liu *et al.*, 2013). These increasingly discovered family and species level patterns of

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press

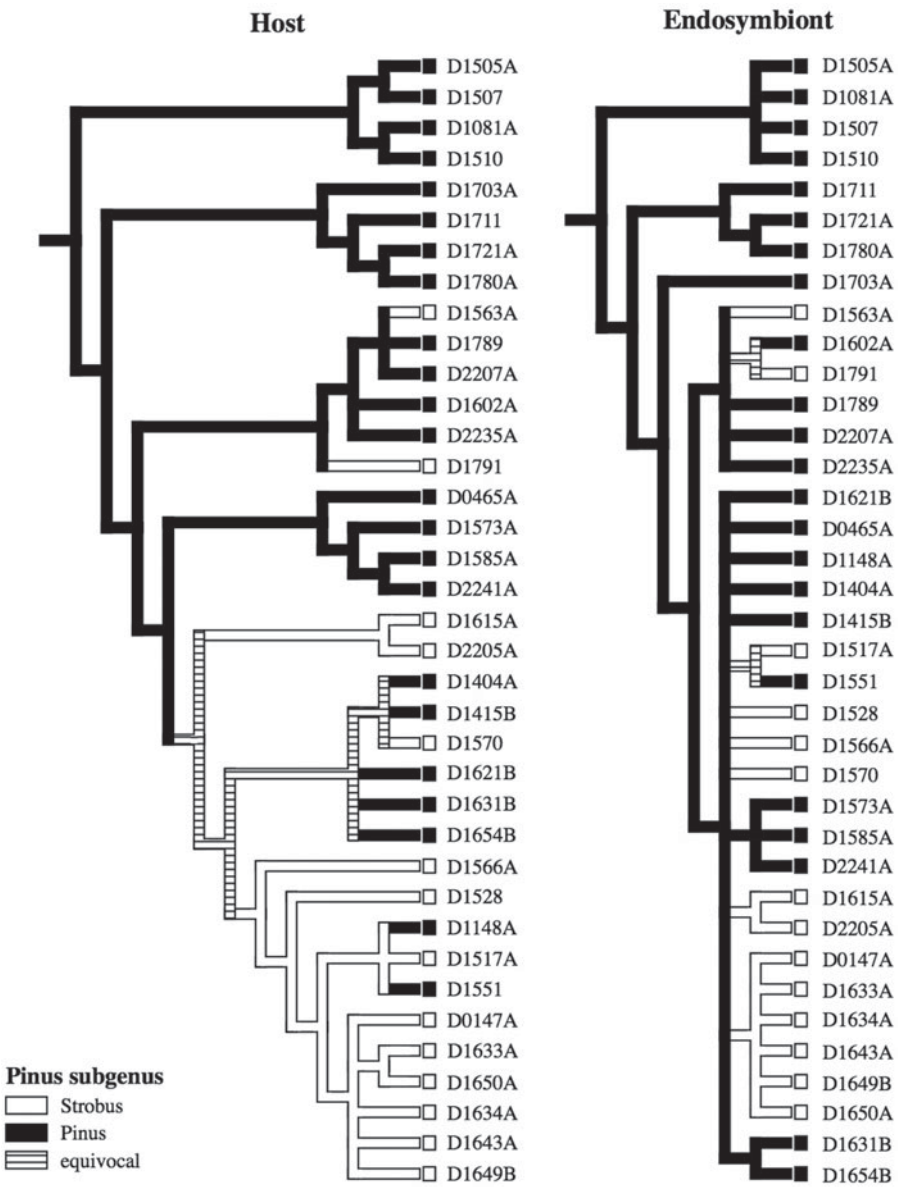

Fig. 4. Plant host association patterns based on collection information published in Gwiazdowski *et al.* (2011). Subgenus designations for the species of *Pinus* from which each individual was collected were mapped onto the host (left) and endosymbiont (right) phylogenies using the trace character feature in MacClade v 4.08 (Maddison & Maddison, 2005).

co-diversification support the increased use of endosymbiont sequence data to help reconstruct host phylogenetic patterns (Lozier *et al.*, 2007; Andersen *et al.*, 2010; Bennett & O'Grady, 2012).

When analyzed independently, all of the endosymbiont loci showed patterns of haplotype sharing between at least two species of *Chionaspis*, though in general these results indicated that individual loci were mostly congruent with host species designations. For 16S the sharing of haplotypes was seen between closely related species (species S6, S7, S8 and S10), while for rspB and for GroEL individuals from S1 shared haplotypes with distantly related species (S7, S8 and S10 for rspB, and S6 for GroEL). It is unclear whether differences are the result of incomplete lineage sorting, horizontal transfer, or some other cause, though for all endosymbiont loci sequence diversity was quite low (table 2). Future studies should take advantage of the recently published *U. diaspidicola* genome (Sabree *et al.*, 2013) to target endosymbiont loci with comparable levels of genetic diversity to developed host loci when making species level comparisons for this or other groups of armored scales.

The co-evolutionary arms race between plants and herbivores has long been hypothesized to be one of the driving forces of speciation in insects through the creation of new adaptive zones (Ehrlich & Raven, 1964), though what role endosymbionts play in shaping patterns of plant–host use remains unclear. The recent sequencing of the *U. diaspidicola* genome (Sabree *et al.*, 2013) confirms the importance of this primary endosymbiont for providing essential amino acids and performing nitrogen recycling, and also suggests that *U.*

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press

diaspidicola has a relationship with its armored scale hosts similar to that seen between aphids and *Buchnera aphidicola*. Given this importance, we expected to see phylogenetic patterns between *U. diaspidicola* (and its host species of *Chionaspis*) and the plant hosts in the genus *Pinus* from which it was collected. However, even when we examined plant–hosts classified at the sub-genus level (fig. 4) we detected no discernable plant and insect/endosymbiont associations. Though contrary to our expectations, this lack of associations between the primary endosymbiont and plant–hosts is similar to the recent results of Toju & Fukatsu (2011). Interestingly, though the authors found no plant–host associations with the primary endosymbiont of the chestnut weevil, they did find highly significant associations between several secondary endosymbionts and plant–hosts, suggesting that secondary endosymbionts may play a more important role than primary endosymbionts in the creation of new adaptive zones.

In contrast to the plant–hosts findings, we did find clear evidence of biogeographic associations (fig. 3). These results are similar to those found in aphids (Liu *et al*., 2013) where the geographical patterns were visible in both the host and endosymbiont phylogenies. Our results indicate that species in the *C. pinifoliae–C. heterophyllae* species complex can be divided into three broad geographic regions: eastern United States, including *C. heterophyllae*, all members of species S7, and one member of species S10; Mexico, including all members of species S2 and S4; and western United States and Baja California, including all members of species S6, S8, and all but one individual of species S10. These patterns were nearly identical between host and endosymbiont phylogenies, however they differed in that the endosymbiont phylogeny reconstructed Mexico as the region of origin for species S2–S10, with a single expansion into the western United States and Baja California, followed by at least one subsequent expansion into the eastern United States, whereas the ancestral relationships in the host phylogeny were unresolved.

In conclusion, we found that sequence data from *U. diaspidicola* and from its hosts in the *C. pinifoliae–C. heterophyllae* species complex show patterns of co-diversification similar to those previously observed between *U. diaspidicola* and its hosts at the family-level suggesting that co-diversification among armored scale insects and *U. diaspidicola* may be occurring at the species level.

Acknowledgements

We thank Gord Burke, Jason Cryan, and Julie Urban, and four anonymous reviewers for valuable comments that improved this manuscript. For financial support we wish to thank Dr Robert Light and the Penn State Erie, The Behrend College undergraduate research grants. We also thank. K. Foust and A. Doolittle for their laboratory assistance. M. Gruwell would like to personally thank S. Gruwell for support. The authors declare no conflict of interest.

References

Ahmed, M.Z., De Barro, P.J., Ren, S.-X., Greeff, J.M. & Qiu, B.-L. (2013) Evidence for horizontal transmission of secondary endosymbionts in the *Bemisia tabaci* cryptic species complex. *PLoS ONE* **8**, e53084.

Andersen, J.C., Wu, J., Gruwell, M.E., Gwiazdowski, R.A., Santana, S.E., Feliciano, N.M., Morse, G.E. & Normark, B.B. (2010) A phylogenetic analysis of armored scale insects (Hemiptera: Diaspididae), based upon nuclear, mitochondrial, and endosymbiont gene sequences. *Molecular Phylogenetics and Evolution* **57**, 992–1003.

Andersen, J.C., Gwiazdowski, R. & Gruwell, M.E. (2014) Molecular evolution of sexual and parthenogenetic lineages of the armored scale insect *Aspidiotus nerii* (Hemiptera: Diaspididae) and its primary bacterial endosymbiont, *Uzinura diaspidicola*. *Annals of the Entomological Society of America* **107**, 954–960.

Baumann, P. (2005) Biology of bacteriocyte-associated endosymbiots of plant sap-sucking insects. *Annual Review of Microbiology* **59**, 155–189.

Bennett, G.M. & O'Grady, P.M. (2012) Host-plants shape insect diversity: phylogeny, origin, and species diversity of native Hawaiian leafhoppers (Cicadellidae: Nesophrosyne). *Molecular Phylogenetics and Evolution* **65**, 705–717.

Buchner, P. (1965) *Endosymbiosis of Animals with Plant Microorganisms*. New York, Interscience Publishers, Inc.

Clement, M., Posada, D. & Crandall, K. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657–1660.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. & Wilson, A. (2012) Geneious v 5.6.2, Available online at http://www.geneious.com computer program, version By Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., and Wilson, A.

Earle, C.J. (2014) The Gymnosperm Database, Available online at http://www.conifers.org/zz/gymnosperms.php.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797.

Ehrlich, P.R. & Raven, P.H. (1964) Butterflies and plants – a study in coevolution. *Evolution* **18**, 586–608.

Gruwell, M.E., Morse, G.E. & Normark, B.B. (2007) Phylogenetic congruence of armored scale insects (Hemiptera: Diaspididae) and their primary endosymbionts from the phylum Bacteroidetes. *Molecular Phylogenetics and Evolution* **44**, 267–280.

Gruwell, M.E., Wu, J. & Normark, B.B. (2009) Diversity and phylogeny of Cardinium (Bacteroidetes) in armored scale insects (Hemiptera: Diaspididae). *Annals of the Entomological Society of America* **102**, 1050–1061.

Gruwell, M.E., Hardy, N.B., Gullan, P.J. & Dittmar, K. (2010) Evolutionary relationships among primary endosymbionts of the mealybug subfamily Phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae). *Applied and Environmental Microbiology* **76**, 7521–7525.

Gwiazdowski, R., Vea, I., Andersen, J.C. & Normark, B.B. (2011) Discovery of cryptic species among North American pine-feeding *Chionaspis* scale insects (Hemiptera: Diaspididae). *Biological Journal of the Linnean Society* **104**, 47–62.

Huelsenbeck, J.P. (2001) MRBAYES: bayesian inference of phylogeny, Version 3.0. *Bioinformatics* **17**, 754–755.

Jepson Flora Project (eds). (2014) Jepson eFlora, Available online at http://ucjeps.berkeley.edu/IJM.html.

Kuechler, S.M., Gibbs, G., Burckhardt, D., Dettner, K. & Hartung, V. (2013) Diversity of bacterial endosymbionts and bacteria–host co-evolution in Gondwanan relict moss bugs (Hemiptera: Coleorrhyncha: Peloridiidae). *Environmental Microbiology* **15**, 2031–2042.

Liu, L., Huang, X., Zhang, R., Jiang, L. & Qiao, G. (2013) Phylogenetic congruence between *Mollitrichosiphum* (Aphididae:

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press

Greenideinae) and *Buchnera* indicates insect-bacteria parallel evolution. *Systematic Entomology* **38**, 81–92.

Lozier, J.D., Roderick, G.K. & Mills, N.J. (2007) Genetic evidence from mitochondrial, nuclear, and endosymbiont markers for the evolution of host plant associated species in the aphid genus *Hyalopterus* (Hemiptera : Aphididae). *Evolution* **61**, 1353–1367.

Maddison, D., & Maddison, W. (2005) MacClade 4 computer program, version 4.08. By Maddison, D., and Maddison, W., Sunderland, MA.

Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. pp. 1–8 in *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA.

Moran, N.A. (2001) The coevolution of bacterial endosymbionts and phloem-feeding insects. *Annals of the Missouri Botanical Garden* **88**, 35–44.

Moran, N.A., McCutcheon, J.P. & Nakabachi, A. (2008) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* **42**, 165–190.

Moulton, J. & Wiegmann, B. (2004) Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta). *Molecular Phylogenetics and Evolution* **31**, 363–378.

Posada, D. & Crandall, K. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Rambaut, A., & Drummond, A. (2007) Tracer v1.4, Available online at http://beast.bio.ed.ac.uk/Tracer computer program, version By Rambaut, A., and Drummond, A.

Sabree, Z.L., Huang, C.Y., Okusu, A., Moran, N.A. & Normark, B.B. (2013) The nutrient supplying capabilities of *Uzinura*, an endosymbiont of armoured scale insects. *Environmental Microbiology* **15**, 1988–1999.

Swofford, D.L. (2002) PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods) computer program, version 4.0b10. By Swofford, D.L., Sunderland, MA.

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.

Urban, J. & Cryan, J. (2012) Two ancient bacterial endosymbionts have coevolved with the planthoppers (Insecta: Hemiptera: Fulgoroidea). *BMC Evolutionary Biology* **12**, 87.

Vea, I., Gwiazdowski, R. & Normark, B.B. (2012) Corroborating molecular species discovery: four new pine-feeding species of *Chionaspis* (Hemiptera, Diaspididae). *Zookeys* **270**, 37–58.

von Dohlen, C.D., Kohler, S., Alsop, S.T. & McManus, W.R. (2001) Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature* **412**, 433–436.

Weisburg, W.G., Barns, S.M., Pelletier, D.A. & Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology* **173**, 697–703.

https://doi.org/10.1017/S0007485314000820 Published online by Cambridge University Press