MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 5751–5766 doi: 10.1111/mec.13423

The origin and genetic differentiation of the socially parasitic aphid *Tamalia inquilinus*

DONALD G. MILLER III,* SARAH P. LAWSON,†[1] DAVID C. RINKER,† HEATHER ESTBY† and PATRICK ABBOT†

*Department of Biological Sciences and Center for Water and the Environment, California State University, Chico, CA 95929, USA, †Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA

Abstract

Social and brood parasitisms are nonconsumptive forms of parasitism involving the exploitation of the colonies or nests of a host. Such parasites are often related to their hosts and may evolve in various ecological contexts, causing evolutionary constraints and opportunities for both parasites and their hosts. In extreme cases, patterns of diversification between social parasites and their hosts can be coupled, such that diversity of one is correlated with or even shapes the diversity of the other. Aphids in the genus *Tamalia* induce galls on North American manzanita (*Arctostaphylos*) and related shrubs (Arbutoideae) and are parasitized by nongalling social parasites or inquilines in the same genus. We used RNA sequencing to identify and generate new gene sequences for *Tamalia* and performed maximum-likelihood, Bayesian and phylogeographic analyses to reconstruct the origins and patterns of diversity and host-associated differentiation in the genus. Our results indicate that the *Tamalia* inquilines are monophyletic and closely related to their gall-forming hosts on *Arctostaphylos*, supporting a previously proposed scenario for origins of these parasitic aphids. Unexpectedly, population structure and host-plant-associated differentiation were greater in the non-gall-inducing parasites than in their gall-inducing hosts. RNA-seq indicated contrasting patterns of gene expression between host aphids and parasites, and perhaps functional differences in host-plant relationships. Our results suggest a mode of speciation in which host plants drive within-guild diversification in insect hosts and their parasites. Shared host plants may be sufficient to promote the ecological diversification of a network of phytophagous insects and their parasites, as exemplified by *Tamalia* aphids.

Keywords: AMOVA, *Arctostaphylos*, Emery's rule, gall, host-associated differentiation, inquiline, RNA-seq, social parasite, trophic interaction

Received 7 November 2014; revision received 7 October 2015; accepted 9 October 2015

Introduction

The mode and tempo of biological diversification emerge from a multitude of ecological and evolutionary forces acting on living systems (Dieckmann *et al.* 2012; Nosil 2012). One recurring theme is how diversity drives further diversity: do species create opportunities for speciation (MacArthur 1972; Strong *et al.* 1984; Farrell *et al.* 1992; Emerson & Kolm 2005; Janz *et al.* 2006)? In phytophagous insects and their host plants, for example, evolutionary radiation on plants can generate host races of specialist herbivores (host associated diversification), which in turn provides new resource opportunities for natural enemies of herbivores, potentially generating cascading or sequential patterns of diversification that extend across multiple interacting species occupying different trophic levels (Abrahamson *et al.* 2003; Stireman *et al.* 2005; Craig *et al.* 2007; Abrahamson & Blair 2008; Evans *et al.* 2013). Ecological theories of speciation have

Correspondence: Patrick Abbot, Fax: (615) 343 6707;
E-mail: patrick.abbot@vanderbilt.edu
[1]Present address: Department of Biological Sciences, University of New Hampshire, Durham, NH 03824, USA

focused on the role of intraspecific competition in fostering divergent selection on populations distributed along a gradient of resources or habitats (Schluter 2000). Renewed attention on the ecological drivers of speciation has highlighted the need to account for the role of interactions in ecological communities, apart from competition (e.g. predation, parasitism, and mutualism; Langerhans 2007; Dieckmann *et al.* 2012). While empirical studies remain relatively rare (Holt & Lawton 1994; Abrams 2000; Karvonen & Seehausen 2012), recent efforts have focused on diversification as community-level ecological and co-evolutionary processes, while explicitly incorporating multiple types of species interactions that span populations to species (Abrahamson & Blair 2008; Feder & Forbes 2010; Violle *et al.* 2012; Althoff *et al.* 2014; Hembry *et al.* 2014).

In this study, we explored patterns of diversification in a group of gall-inducing insects involved in an unusual form of brood parasitism on their host plants. Gall-forming insects have provided some of the best examples of multi-trophic diversification, presumably because they are characterized by highly specialized interactions between herbivores, plants and often natural enemies, and are amenable to both ecological and evolutionary studies (Abrahamson *et al.* 2003; Stireman *et al.* 2010). Aphids in the genus *Tamalia* are herbivores on New World Arbutoideae (Ericales: Ericaceae) in western North America. *Tamalia* foundress females singly, or in groups, occupy hollow, tumour-like galls that they induce on the leaves of their host plants (Miller 1998a). These galls are invaded by a closely related congener that has lost the capacity to induce galls, and that depends instead upon its gall-forming hosts for persistence. Because invading, non-gall-forming *Tamalia* diminish resources and space within galls that would otherwise be available to host aphids, the interaction is parasitic (Miller 2004). The interaction between *Tamalia* gall formers and their parasites (also described as *inquilines*) closely resembles other forms of nonconsumptive social or brood parasitism. Brood parasitism occurs when nesting behaviours of birds and insects are appropriated by heterospecific females that 'dump' their eggs into the nests of foster parents (Davies & Brooke 1988; Zink 2000; Stoddard & Kilner 2013). Social parasitism occurs in some ants and other Hymenoptera, in which parasitic species utilize the brood care behaviour of host workers, and results in the production of a parasitic worker brood or sexuals at the expense of host ant brood (Buschinger 1986). Brood and social parasites and their hosts often exhibit striking patterns of diversification (Litman *et al.* 2013), and, in some cases, parasitic species originate from and attack their nearest phylogenetic relatives, an evolutionary syndrome known as the strict form of Emery's Rule (Emery 1909; Lowe *et al.* 2002; Smith *et al.* 2007). The evidence for and against Emery's Rule has been equivocal (Huang & Dornhaus 2008), and it has scarcely been investigated in gall-inducing taxa (Ronquist 1994; Miller & Crespi 2003; Gilbert *et al.* 2012). Previous work suggested that both hosts and parasitic aphids are radiating across New World manzanitas and relatives (Miller & Crespi 2003). Gall-inducing *Tamalia* aphids thus offer novel opportunities to study patterns of diversification in plant herbivores and their brood or social parasites.

Our goals were to characterize the timing and possible circumstances of the origins of social parasitism in *Tamalia* aphids, and the patterns of host-plant-associated diversification in host and parasitic aphids. Gall-inducing insects tend to be host-plant specialists, presumably because of the life history and biochemical constraints associated with forming galls on plants (Wool 2004). Because it is likely that parasitic aphids experience relatively few such constraints, we expected that they would express less specificity across host plants than their gall-forming host aphids. If so, this would suggest that the diversification of host and parasitic aphids on manzanita and madrone host plants is relatively decoupled, despite their obligate, parasitic interaction. We sampled *Tamalia* gall inducers and inquilines broadly throughout their range on a diverse set of 13 host-plant species in the western United States. We used Illumina sequencing to develop genes for phylogenetic reconstruction from nuclear and bacterial symbiont genomes, as well as to explore evidence for host-associated differentiation in *Tamalia* gall inducers and inquilines in the light of the predictions of Emery's Rule, viz., that social parasites and their hosts are each others' closest phylogenetic relatives. Our RNA-seq data also offered opportunities to describe possible differences in gene expression related to the gall-forming and non-gall-forming lifestyles. Both inquilines and their gall-forming hosts are monophyletic and appear to be radiating on manzanita (*Arctostaphylos*) shrubs, and our results provide moderate although not unequivocal support for the hypothesis that inquilines evolved sympatrically with their hosts on *Arctostaphylos*, without an intervening host-plant shift. Surprisingly, we found a relatively greater divergence along host-plant lines in inquilines than in gall formers. Because of the population-genetic consequences of the parasitic lifestyle, sorting of genetic variation across host plants appears to occur faster in inquilines than in their gall-forming hosts that they by necessity track, resulting in a pattern of seemingly synchronous diversification.

Materials and methods

Study system

A gall is the result of intricate interactions between insect and host plant; consequently, gall formers are typically host-plant specialists (Wool 2004). Galls provide essential food, shelter and space for reproduction in insects. Those insect taxa inducing hollow (rather than solid) galls, such as thrips and aphids, may exhibit social behaviours, ranging from multiple-foundress associations to societies defended by a sterile soldier caste (Foster 2002; Chapman *et al*. 2008). The galls themselves are often subject to invaders or usurpers (e.g. Akimoto 1988, 1989).

Tamalia is a genus of Nearctic aphids in the subfamily Tamaliinae, consisting of five described (and three undescribed) species that form galls in western North America on manzanitas, madrones and summer holly (*Arctostaphylos*, *Arbutus* and *Comarostaphylis*; Ericales: Ericaceae: Arbutoideae) (Remaudière & Stroyan 1984; Blackman & Eastop 2006). Taxonomically, *Tamalia* has been of interest because it retains several morphological traits that are unusual in aphids, including sexually reproducing females with wings (Remaudière & Stroyan 1984). In the gall inducers, as many as 10 cofoundress females (e.g. *Tamalia coweni*) may share gall space without agonistic interactions (Miller 1998a). The galls of each are parasitized by the recently described inquiline *Tamalia inquilinus* (Miller & Sharkey 2000), which invades the galls of its host and reduces the reproductive output of its host gall inducers (Miller 2004). Experimental and circumstantial evidence indicates that *Tamalia* inquilines cannot initiate galls on their own; hence, they are obliged to locate and enter existing galls, typically in the early or late stages of formation (Hamilton 1987; Miller 2004). Previous work suggested that the parasitic inquiline clade may have radiated in parallel with its host plants and host aphids (Miller & Crespi 2003). Before ecological details of the interaction were better understood, Miller & Crespi (2003) treated inquilinism (which involves no harm to the host) as distinct from parasitism (which does cause harm to the host). Subsequent experimental evidence has revealed that the inquilines are indeed parasites (Miller 2004). Communal gall occupation by *Tamalia* foundresses incurs a *per capita* loss in survival and reproduction to foundresses, yet overall group fitness may be increased, because of high mean relatedness among foundresses (Miller 1998b; Taylor & Miller 2014). Thus, facultative communal occupation (and mutual tolerance of cofoundresses) within galls can be hypothesized as a mechanism facilitating the exploitation of galls by conspecifics, culminating in the evolution of obligate, heterospecific inquiline specialists (Miller 2005). Assuming inquilinism arose from a gall-inducing ancestor, the loss of gall-inducing ability is secondarily derived. In all populations of both gall inducers and inquilines examined, communal behaviour has been documented, suggesting it is plesiomorphic (Miller 2005).

Sample collection

Tamalia aphid galls were collected from 13 species of host plants at 17 sites across Arizona, Nevada and California (Fig. 1; Tables S1 and S2, Supporting information). A total of six *Tamalia* species were collected: *T. coweni* (Cockerell), *Tamalia dicksoni* Remaudière & Stroyan, *T. inquilinus* Miller, and three novel entities, *Tamalia* Species A, *Tamalia* Species B and *Tamalia* Species C. All collections were made between May 2010 and May 2011. Aphids were dissected from fresh galls and preserved immediately at −80 °C. Subsequently, aphids within 1.5-mL tubes were immersed in liquid nitrogen and pulverized with a micropestle prior to DNA or RNA extractions. Voucher specimens will be deposited in the Smithsonian Institution Department of

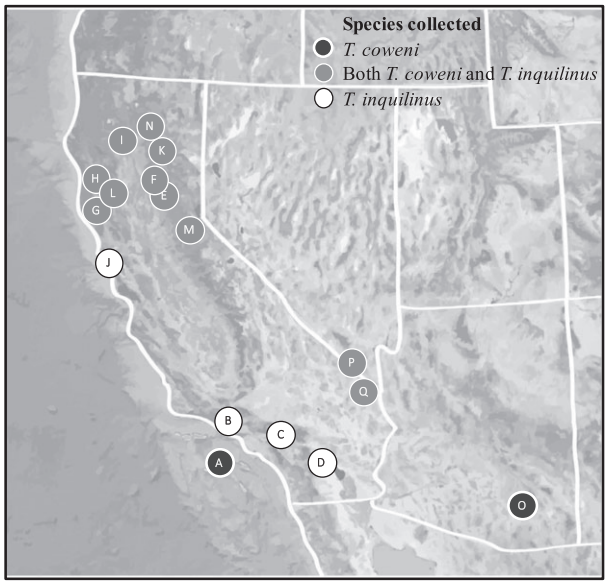

Fig. 1 Sampling locations of *Tamalia* species used in this study. A total of six *Tamalia* species were sampled. The shading of the circle (black, white or grey) highlights the sampling locations of the two focal species for the population-level analysis on host-associated diversification between a gall former and a parasite (*Tamalia coweni* and *Tamalia inquilinus*). Black = *T. coweni*; white = *T. inquilinus*; grey = both *T. coweni* and *T. inquilinus*. The lettered circles correspond to sample locations for all six *Tamalia* species collected and are described in detail in Tables S1 and S2 (Supporting information). *Tamalia dicksoni* collected at sites: B, D, G, J; *Tamalia* species A: O; *Tamalia* species B: A; *Tamalia* species C: B, C, D.

Entomology, Canadian National Collection of Insects, Washington State University and California State University, Chico.

High-throughput sequencing

Total genomic RNA was extracted from four *Tamalia* samples of 50 adult animals each, representing *T. coweni* and *T. inquilinus* populations on the host-plants *Arctostaphylos viscida* and *Arctostaphylos manzanita* in northern California. Aphid samples were preserved in an ultracold freezer, and separated by age (juveniles and adults) on dry ice. Adults were pooled from ~10 galls, and ground in liquid nitrogen. We used the Qiagen RNeasy Mini Kit® for cells, tissues and yeast, including the DNase digestion protocol. Purified RNA was quantified with a NanoDrop ND-1000 spectrophotometer. Four libraries were constructed and sequenced through a paired-end read protocol on the Illumina Genome Analyzer at the HudsonAlpha Institute for Biotechnology Genome Services Lab (hudsonalpha.org). Illumina instrument software was used to perform data analysis and base calling.

Data preprocessing, de novo transcriptome assembly and transcript annotation

Initial quality assessment of raw RNA-seq data, performed using FASTQC (version 0.10.1; bioinformatics.babraham.ac.uk/projects/fastqc/), revealed adapter contamination and low-quality reads. To ensure the quality of transcriptome analyses, quality-based trimming and adapter trimming were performed using Trimmomatic (version 0.22; Lohse *et al.* 2012) with the following parameters: LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15 and MINLEN:25. All data filtrations were conducted in paired-end mode. Cleaned data from all four samples were combined and assembled using SOAPDENOVO-TRANS (version 1.03; Xie *et al.* 2013) to generate a de novo transcriptome assembly. SOAPDENOVO-TRANS was run with default settings, and the longest locus from each scaffold was taken as the representative isoform of that putative transcript.

The nucleotide sequence of each assembled locus was then compared against the annotated consensus sets of mRNA and protein sequences of the pea aphid (Hemiptera: Aphididae: *Acyrthosiphon pisum*; version 2.1; aphidbase.com) using a local BLASTN and BLASTX (Altschul *et al.* 1997) in conjunction with an *e*-value cutoff of 1e-4. In cases where the two search methods produced divergent results, the top pea aphid ortholog from the BLASTX search was taken. *Tamalia* transcripts with identifiable orthologs in the pea aphid genome were then further annotated with Blast2GO predictions for pea aphid genes (version 2.1; http://aphidbase.com).

Read quantitation and differential expression

Reads from each sample were aligned to the longest full loci in the indexed genome using BOWTIE2 (Langmead & Salzberg 2012). Statistical significance along with fold change was determined by pairwise comparison of the BOWTIE2 alignments for pairs of samples using GFOLD (generalized fold change) configured for a 95% confidence interval (Feng *et al.* 2012). The result was a set of GFOLD values (i.e. GFOLD's 'reliable' log2 fold change) for each assembled locus. GFOLD values >0 were considered significantly, differentially expressed. Reads mapping to each locus were also extracted from the BOWTIE2 alignments using the read count option in GFOLD.

Gene ontology enrichment

The BiNGO plugin (Maere *et al.* 2005) for CYTOSCAPE (v.3.1.0; Shannon *et al.* 2003) was used to evaluate Gene Ontology (GO) terms enriched for selected subsets of assembled loci. Enrichment was determined by a hypergeometric test using a false discovery rate procedure at a significance level of 0.05 (Benjamini & Hochberg 1995). Patterns of overlap between assembled loci across treatments were visualized with VENNY 2.02 (Oliveros 2007–2015).

Primer design, DNA amplification and sequencing

There are currently no *Tamalia* nuclear gene sequences in GenBank. In order to develop genetic markers for *Tamalia*, we used OrthoDB (http://orthodb.org/orthodb7) to identify single-copy orthologs in the pea aphid and used these gene sequences to interrogate the *Tamalia* RNA-seq data set for putative single-copy candidates for phylogenetic analysis. We selected candidates that exhibited 10–30% divergence between the pea aphid and *Tamalia*, reasoning that divergence at the species level may be a correlate of sequence diversity within *Tamalia*. Using pea aphid gene models, we designed degenerate primers in neighbouring exons where possible. After screening, we selected three nuclear aphid genes with coding sequences, one intronic sequence and one gene from the aphid symbiont *Buchnera* (Table S3, Supporting information): an ortholog of double-stranded RNA-specific editase (DME/*adar*); an ortholog of 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II (DME/*Plc21c*); an ortholog of lysine-specific demethylase lid (DME/*lid*); a putative intron of phosphoenolpyruvate carboxykinase (DME/*PEPCK*); and the

Buchnera groEL. Symbiont-derived genes from *Buchnera* have proven useful in the reconstruction of aphid phylogenetic relationships (Clark *et al.* 2000; Nováková *et al.* 2013). Total genomic DNA was extracted from aphid specimens using the Qiagen DNeasy Blood and Tissue Kit®. Degenerate primers were used to amplify and sequence segments of the three nuclear genes: PCR products were purified using ExoSAP-IT® (USB Corporation, Cleveland, OH, USA) and then Sanger-sequenced at the Vanderbilt Technologies for Advanced Genomics core (vantage. vanderbilt.edu) or GENEWIZ, Inc. (genewiz.-com).

Phylogenetic analysis

Sequences were edited and aligned in GENEIOUS version R7.1 (Biomatters; http://www.geneious.com), using the MAFFT v6.814b plug-in program in the GENEIOUS software (Katoh *et al.* 2002). The alignment strategy was determined automatically in the MAFFT/GENEIOUS implementation, and we used a scoring matrix of 200PAM/k = 2 and a gap open penalty of 1.53. Subsequent alignments were manually curated, and beginning and ending gaps were removed. Maximum-likelihood (ML) and Bayesian methods were used in the phylogenetic analysis, on 82 samples from six *Tamalia* species. We performed two categories of analysis. First, we analysed each gene tree individually, rooting the trees when possible with publically available aphid sequences from GenBank. For the nuclear genes, rooting was done with *A. pisum* sequences (XM_001948842, XM_008185038 and XM_008187279). For the *Buchnera* groEL gene tree, more closely related sequences from aphids in the subfamily Thelaxinae were available from a study by Nováková *et al.* (2013) (*Glyphina longisetosa* (JQ269418); *Kurisakia querciphila* (JQ269419); and *Thelaxes californica* (KC213246)). No alignable sequences were available for the PEPCK intron. We suspected that the distantly related *A. pisum* would result in tenuous rooting of *Tamalia* gene trees, and tested the reliability of alternative roots suggested by the nuclear genes by conducting Shimodaira approximately unbiased (AU) tests on alternative topologies (Shimodaira & Hasegawa 1999; Shimodaira 2002). Trees with alternative roots were constructed in MESQUITE v. 3.03 (Maddison & Maddison 2015), and AU tests were conducted in PAUP* ver. 4.0a146 using the RELL option, codon partitions (where possible), a GTR rate matrix, gamma distributed rates estimated from the data, and 10 000 bootstrap replicates (Swofford 2002). Secondly, we analysed a concatenated matrix consisting of the three nuclear genes without outgroup rooting (retaining the Thelaxinae-rooted *Buchnera* groEL matrix separately). Concatenated matrices were created in MESQUITE. For all phylogenetic analyses, matrices were partitioned by genes, codons and introns (when present), and best nucleotide substitution models for combinations of gene partitions were determined in Partition Finder ver 1.1.1 (Lanfear *et al.* 2012), using the 'greedy' algorithm, linked branch lengths, and evolutionary models and model selection (AIC or BIC) appropriate to the method of phylogenetic inference. ML analyses were conducted in RAXML 8.1.2 and the RAXMLGUI software (Stamatakis 2006; Silvestro 2012). Using model settings recommended by Partition Finder, we conducted full ML searches, generating the best tree out of 100 independent runs with different starting trees, and bootstrap support values with 1000 replicates using the 'bootstrap and consensus' option. Bayesian trees were inferred with MRBAYES ver 3.2.4 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003), generated from matrices with partitions and models recommended by Partition Finder, with priors set to default values and unlinked, and the rate parameter set to 'variable' across partitions. Two independent runs were performed, each consisting of two parallel Markov chain Monte Carlo runs for 10 million generations, which were sampled every 1000 generations. We inspected the trace files in Geneious and discarded the first 25% of trees. We constructed the majority-rule consensus tree using the remaining trees, with clade support indicated by the Bayesian posterior probabilities. Models and partitions are described in Table S4 (Supporting information). Finally, we tested the taxonomic hypotheses of monophyly for each species with more than two samples using the Shimodaira-Hasegawa (SH) test, as implemented in PAUP* ver. 4.0a146 (Shimodaira & Hasegawa 1999; Swofford 2002). For computational purposes, matrices were first randomly pruned by 30%, and constraint trees to test for monophyly were created in MESQUITE v. 3.03 (Maddison & Maddison 2015).

Intraspecific diversity across plants and populations

We used two separate analyses to quantify the extent to which host-plant associations explain patterns of population-genetic variation in *Tamalia* gall-inducer and inquiline lineages. In the first, we conducted analyses of molecular variance of our sequence data within and among host aphid species, using the GENALEX version 6.5 add-in package for Microsoft Excel (AMOVA; Excoffier *et al.* 1992). *Tamalia coweni* samples from five *Arctostaphylos* host-plant species (*A. viscida*, *A. manzanita*, *Arctostaphylos patula*, *Arctostaphylos canescens* and *Arctostaphylos pungens*) and *T. inquilinus* samples from eight *Arctostaphylos* species (*A. viscida*, *A. manzanita*,

A. patula, *Arctostaphylos pringlei*, *Arctostaphylos glandulosa*, *Arctostaphylos glauca*, *Arctostaphylos nevadensis* and *Arctostaphylos mewukka*) were analysed separately for genetic variance based on the levels of the individual genotype (within populations) and host-plant species (among populations). Because inspection of the species trees generated in the phylogenetic analyses indicated that *T. coweni* and *T. dicksoni* are unlikely to be monophyletic taxa, we also performed an AMOVA as described above, but with *T. dicksoni* included with the *T. coweni* samples. In the second analysis, we accounted for the possibility that patterns of differentiation across host plants were confounded by geographic subdivision by repeating the AMOVA described above, but analysing only the *T. coweni* and *T. inquilinus* sampled from the same sites and host-plant species (*A. viscida*, *A. manzanita* and *A. patula*). The latter analysis, in effect, measures the *relative* difference between *T. coweni* and *T. inquilinus* in how genetic variation is partitioned across host plants and geography, but does not resolve the potential confounding effect of geographical subdivision on host-associated differentiation. To evaluate the contrasting effects of host plant and geography on genetic differentiation, we conducted Mantel and partial Mantel tests on genetic, geographic and host-plant distance matrices (Smouse *et al.* 1986). We conducted the analysis separately for three genes with the most sequence diversity (the PEPCK intron, the nuclear Plc21c and the *Buchnera* groEL) and constructed haplotype networks for the latter two in order to visualize the patterns of differentiation across host plants. Matrices of genetic distances were generated in MESQUITE, using uncorrected *p* distances and converting ambiguities to missing data. Geographic distance matrices in units of kilometres were generated in GEOGRAPHICDISTANCEMATRIX v.1.2.3, using the default WGS84 radius reference system (Ersts 2015). Following Drummond *et al.* (2009), distance matrices for host-plant use were generated by coding shared hosts as '0' and distinct hosts as '1'. We then performed haplotype network reconstructions, in order to visually summarize the patterns of subdivision across host plants and geography. We first determined phase of Plc21c using PHASER 2.1 (Stephens *et al.* 2001). We then constructed optimal parsimony networks in TCS v.1.21 (Clement *et al.* 2000). In the case of *T. inquilinus* groEL, divergence between haplotypes was large, and we generated a median-joining network in NETWORK v.4.613 (Bandelt *et al.* 1999), the subnetworks of which were congruent to those generated in TCS. We generated comparative summary statistics on sequence polymorphism and haplotype diversity across host plants and geography for *T. coweni* and *T. inquilinus* in DNASP v.5.10.1 (Librado & Rozas 2009).

Results

Phylogenetic analysis

Maximum-likelihood and Bayesian analysis of the unrooted 1548 bp concatenated nuclear gene alignment recovered a topology with support for the monophyly of most *Tamalia* species, with the exception of *Tamalia coweni* and *Tamalia dicksoni*, which were not resolved (Figs 2 and S1–S4, Supporting information). SH tests supported the monophyly of *Tamalia inquilinus* over minimally altered topologies (e.g., $-\ln L = 6419.3$; difference in $-\ln L$ between inquiline monophyly and the best alternative = 317.8; $P < 0.001$), but not the monophyly of *T. coweni* and *T. dicksoni* vs. their consolidation ($-\ln L = 6708.3$; difference in $-\ln L$ between *T. coweni* + *T. dicksoni* vs. their separate monophyly = 109.3; $P < 0.001$). One sample (*T. coweni* 043) may represent either an undescribed *Tamalia* lineage, or possibly a mislabelled *Tamalia* Species C sample. As expected, outgroup rooting with *Acyrthosiphon pisum* was equivocal and differed between gene trees. Two of the three nuclear gene trees placed the root at *Tamalia* Species A, a gall-forming aphid on *Arbutus*. However, a third (*adar*) placed the root at *T. inquilinus* on *Arctostaphylos*. AU tests provided significant support for a root at *Tamalia* Species A rather than *T. inquilinus* in one case (*lid*; difference in $-\ln L = 7.675$; $P = 0.022$), no significant difference between a *Tamalia* Species A and *T. inquilinus* root (*Plc21c*; difference in $-\ln L = 2.9$; $P = 0.18$), and significant support for a *T. inquilinus* root (difference in $-\ln L = 7.4$; $P = 0.02$). However, both ML and Bayesian reconstruction of the fast-evolving *Buchnera* groEL, rooted with the three Thelaxinae species, placed the root at *Tamalia* Species A on *Arbutus* (Figs S1 and S2, Supporting information). Thus, three of four gene trees provide some support for the scenario derived from an analysis of a mitochondrial gene in Miller & Crespi (2003) that argued for the early divergence of *Tamalia* Species A on *Arbutus*, with the subsequent radiation of *Tamalia* Species B on *Comarostaphylis* and *T. coweni*, *T. dicksoni*, *Tamalia* Species C and *T. inquilinus* on *Arctostaphylos*. The sister taxa relationship between *Tamalia* Species B and the *Arctostaphylos Tamalia* spp., supported in all of our rooted trees other than *adar*, is consistent with the scenario of Miller & Crespi (2003).

In general, we found greater resolution of nodes within the lineage of the socially parasitic species, *T. inquilinus*, than that of gall-forming hosts (Figs 2 and S1–S4, Supporting information), and the topology of the nuclear tree suggests that *T. inquilinus* exhibits at least as much, if not more, structure across host plants than do gall formers (Fig. 2). For example, a group of inquilines collected from two distinct populations (samples 14,

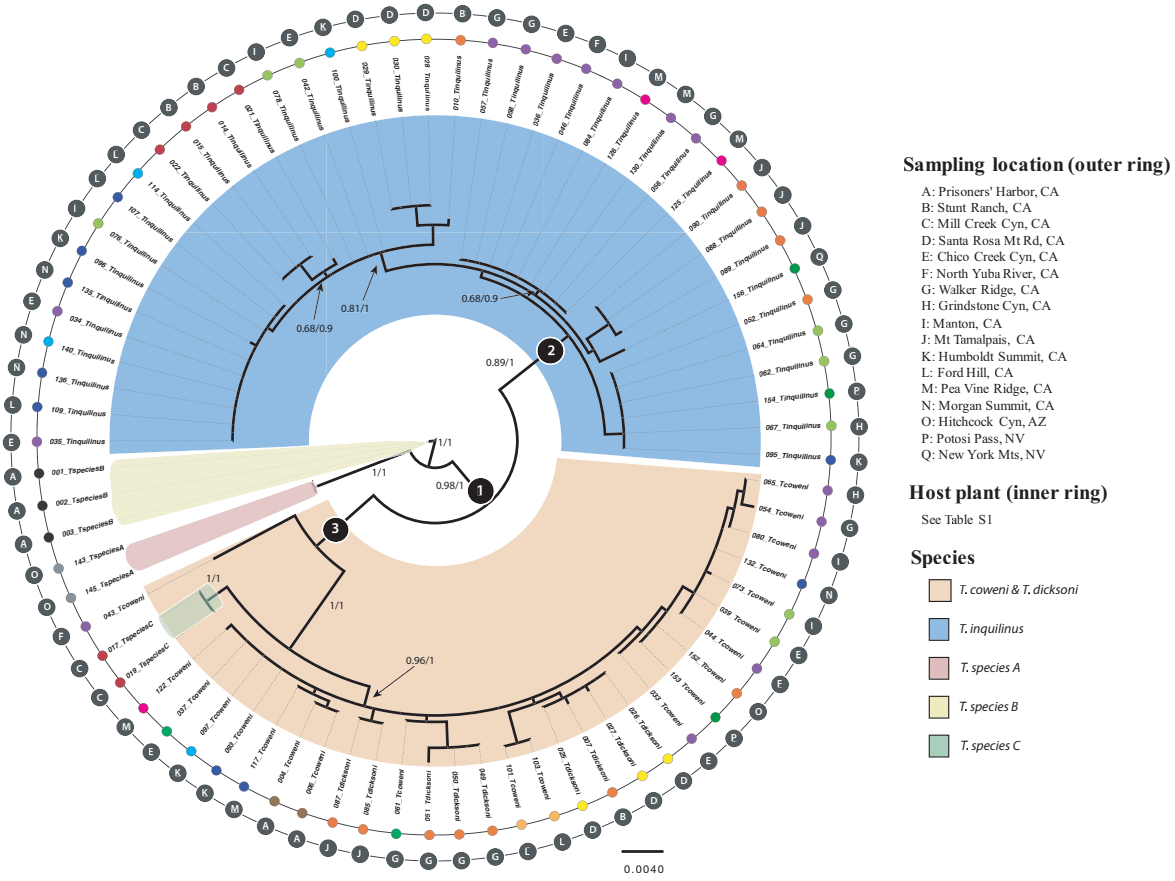

Fig. 2 Three gene maximum-likelihood (ML) tree of *Tamalia* on North America Arbutoideae (midpoint root not shown). The sample identification number for each individual is indicated and corresponds to those itemized in Table S1 (Supporting information). The outer ring indicates the sample locations mapped in Fig. 1 and Table S1 (Supporting information). The colour of the circle next to each taxon name indicates the host plant from which each was collected (see Table S1, Supporting information). ML bootstrap support followed by posterior probabilities from the Bayesian analysis is shown for represented groups. Three partitions are noted, indicated by black circles: (1) the divergence between the gall formers and inquilines on *Arctostaphylos*, and *Tamalia* species B and *Tamalia* species A on *Comarostaphylis* and *Arbutus*, respectively; (2) the split that defines *Tamalia inquilinus*; and (3) the split that defines the gall formers on *Arctostaphylos* (*Tamalia* species C, *Tamalia dicksoni* and *Tamalia coweni*. Inset: corresponding phylogram, with colours matching the circle tree.

15, 21, 22) was resolved in both ML and Bayesian analyses parasitizing *Tamalia* Species C on *Arctostaphylos glauca* (samples 17, 19). The inquiline samples from host plants *Arctostaphylos viscida* and *Arctostaphylos mewukka*, also from multiple populations and the latter thought to be of hybrid origin between *A. viscida* and *Arctostaphylos patula* (Roof 1967; Schierenbeck *et al.* 1992; samples 36, 46, 56, 57, 58, 84, 125, 126, and 130), form a clade with moderate support, as do some of the corresponding samples of the gall-inducing *T. coweni* on the same host plants. Although sampled from only single populations, several other inquiline clades are supported: for example, a cluster of inquilines collected in association with northern California populations of *T. dicksoni* on *Arctostaphylos glandulosa* (samples 88, 89, 90), and inquilines sampled from southern California populations of *T. dicksoni* on *Arctostaphylos pringlei* (samples 28, 29, 30). These host-associated groups are largely mirrored in the *Buchnera* groEL trees (Figs S1 and S2, Supporting information). Sequences have been deposited in GenBank with the Accession nos (KM604232–KM604313; KM604314–KM604395; KM604396–KM604477; KM604478–KM604559; KM604560–KM604641).

Host-associated differentiation

Phylogenetic analyses provide evidence of intraspecific host-plant differentiation in *Tamalia* spp. For both *T. coweni* and *T. inquilinus*, we found a significant degree of differentiation across host plants (among populations; $P < 0.001$ for both *T. coweni* and *T. inquilinus*; Figs 2, S6 and S7, Supporting information; Table 1).

Haplotype networks clearly differ between the gall-forming *T. coweni* and the parasitic *T. inquilinus*, with the latter exhibiting reduced haplotype diversity (Tables S8 and S9, Supporting information) and greater evidence of geographically widespread, host-plant-restricted haplotypes (Figs 3 and S5, Supporting information; see below). The amount of variation explained by host-plant association was greater for the inquiline parasites (49%) than the gall-inducing *T. coweni* hosts, whether or not we analysed *T. coweni* alone or with *T. coweni* and *T. dicksoni* together (32% vs. 34%, respectively). When we confined our analysis to only those sites in which we sampled both *T. coweni* and *T. inquilinus*, the relative differentiation across host plants (*A. viscida*, *A. manzanita* and *A. patula*) remained significant and relatively greater in the parasitic aphid than its host (*T. inquilinus* between host plants = 34% vs. 18% for *T. coweni*; Fig. S7; Table S9, Supporting information). Moreover, partial Mantel tests indicated that there was no evidence of *T. coweni* population structure by geography in Plc21c or the PEPCK intron and that only host-plant association had a significant effect on genetic structure in *T. coweni* PEPCK and groEL, after removing the effect of geographic subdivision (Table 2). However, in the inquiline, there were strong and significant effects of geographic isolation and host-plant association on genetic structure. Partial Mantel tests showed that the pattern of genetic structure is driven almost entirely by host plants. There was no evidence of population structure in Plc21c or PEPCK intron, once host-plant effects were removed. In groEL, population structure explains a significant amount of variation after removing the effect of host plants, but the converse is true as well, and thus for the most variable gene, geographic subdivision and host plant cannot be entirely disentangled (Table 2). However, overall, geographic subdivision explains comparatively less of the genetic variation than does host-plant effect in either host or inquiline aphid.

RNA sequencing and assembly

RNA sequencing of two *T. coweni* and two *T. inquilinus* samples across two host plants produced a total of 115 200 000 high-quality reads and a total of 5.8 billion base pairs (bp) of sequence, with a range of 12–23.6 million reads per sample. Quality-filtered reads were assembled de novo into 121 822 scaffolds overall, with a mean scaffold size of 577 bp, and a N50 of 1786 bp (Table S5, Supporting information). After filtering isoforms, 21 826 unique *Tamalia* loci were generated (mean size = 1416 bp, N50 = 2021). Not surprisingly, given the evolutionary distance between *Tamalia* and the pea aphid, and the large fraction of the pea aphid genome that still lacks functional annotation, BLASTN and BLASTX searches of these loci against *A. pisum* (ACYPI) nucleotide and protein databases produced a modest number of significant hits ($e < 0.0001$): 6500 protein and 12 679 nucleotide hits, and 6287 loci overall annotated with both significant ACYPI nucleotide and protein sequences.

Thus, while the experimental design permitted comparisons of differential expression between *T. coweni and T. inquilinus* adults across two host plants, many of these genes lacked functional annotations. On *A. manzanita* and *A. viscida*, 103 and 149 genes, respectively, were significantly differentially expressed between adult *T. coweni* and *T. inquilinus* after correcting for multiple comparisons, 49 genes of which overlapped across host plants. Overall, there was remarkable consistency in patterns of gene expression across the two host plants. Of the 49 genes, all were differentially expressed in the same direction (*T. coweni* vs. *T. inquilinus*; Spearman's $P = 0.91$). However, only nine had significant BLAST hits to functionally annotated ACYPI genes. Most of these do not have any clear relation to the life history differences of the two aphids. GO ontology searches of *Drosophila* homologs returned activities such as protein and transcription factor binding, and

Table 1 AMOVA table for *Tamalia coweni* and *Tamalia inquilinus* collected across the full set of sampled host plants

	Source	d.f.	SS	MS	Est. var.	%	Value	*P*
T. coweni	Among plants	4	40.833	10.208	2.031	32		
	Within plants	10	42.500	4.250	4.250	68		
	Total	14	83.333		6.281	100		
							0.323	<0.001
T. inquilinus	Among plants	7	183.108	26.158	6.549	49		
	Within plants	16	109.183	6.824	6.824	51		
	Total	23	292.292		13.373	100		
							0.490	<0.001

Codes: d.f., degrees of freedom; SS, sum of squares; MS, mean square; Est. var., estimated variance, *P*, *P*-value from 9999 permutations.

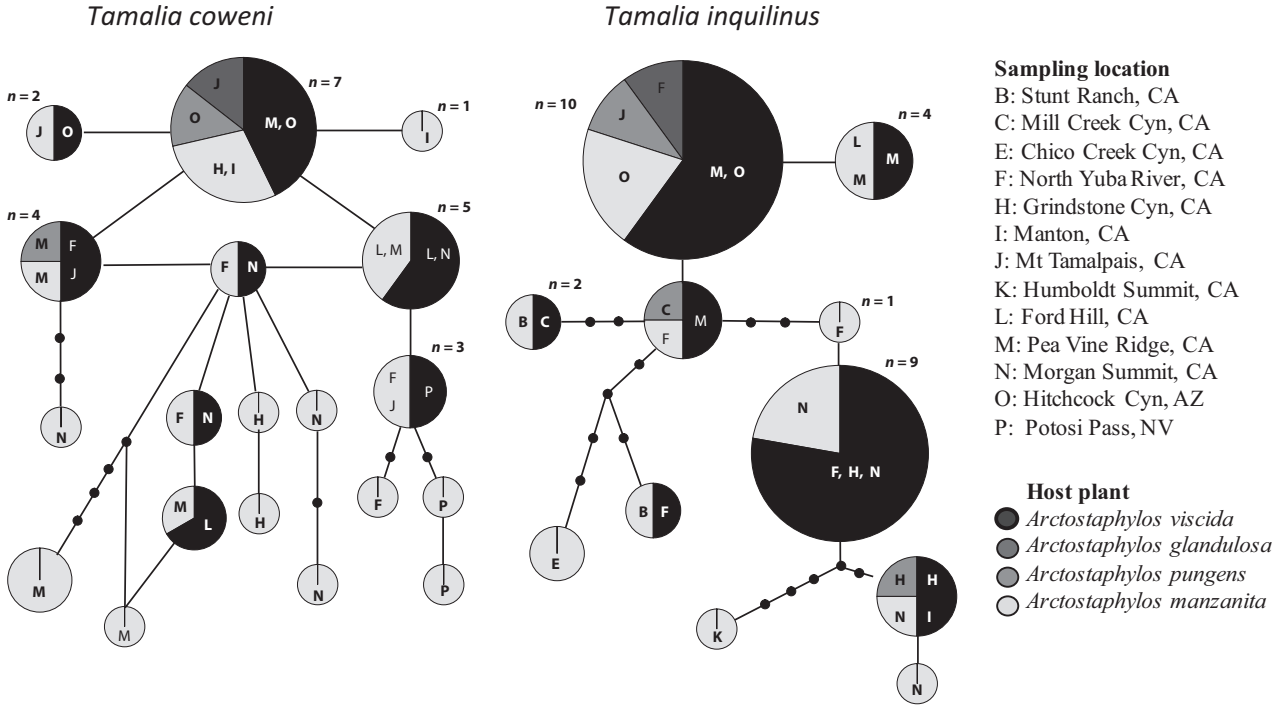

Fig. 3 Haplotype networks for *Tamalia coweni* and *Tamalia inquilinus*, derived from the nuclear *Plc21c* gene. The size of the circles is proportional to the sample size, and representative sample sizes are provided for each. The shading corresponds to a given host plant, and the letters are the geographic sampling locations as provided in Table S2 (Supporting information).

Table 2 Results from Mantel and partial Mantel tests of the correlation between genetic and geographic distances and host-plant association in *Tamalia coweni* and *Tamalia inquilinus* for Plc21c, PEPCK, and groEL

	Pls21c		pepck		groEL	
	r	*P*-value	*r*	*P*-value	*r*	*P*-value
T. coweni						
By population	0.159	—	0.01	—	0.145	—
By host plant	0.07	—	0.16	<0.001	0.190	0.004
By population, partial plants	0.152	—	0.06	—	0.114	—
By plants, partial population	0.053	—	0.17	0.002	0.168	0.015
T. inquilinus						
By population	0.176	0.044	0.17	0.03	0.300	0.009
By host plant	0.29	<0.001	0.34	<0.001	0.238	<0.001
By population, partial plants	0.113	—	0.07	—	0.255	0.022
By plants, partial population	0.186	<0.001	0.30	<0.001	0.173	0.005

various functions related to reproduction. One metabolic gene, upregulated in *T. coweni* (3-hydroxyisobutyrate dehydrogenase-like), possesses oxidoreductase and catabolic activity, and another, upregulated in *T. inquilinus*, encodes a protein with lycopene cyclase/phytoene synthase activity, orthologous to a gene recently identified in the pea aphid genome that was laterally transferred from fungi (Moran & Jarvik 2010). Overall, it is clear that at the level of gene expression, species differences produce far more divergence in gene expression than host-plant associations, as more genes are differentially expressed between *T. coweni* and *T. inquilinus* occupying the same galls on the same host plant than within *T. coweni* across host plants (Fig. 4). The top 100 genes with functional annotations that were significantly differentially expressed between *T. coweni* and *T. inquilinus* on *A. viscida* and *A. manzanita*, respectively, are listed in Tables S6 and S7 (Supporting information). Raw data have been submitted to the NCBI Sequencing Read Archive.

Discussion

Gall-inducing insects can have important effects on the structure of communities in which they are common, because they alter host-plant architecture and provide food or shelter for the predators, parasites and co-existing herbivores and microbes with which they interact (Bailey & Whitham 2003). They have also provided some of the clearest examples of adaptive radiations

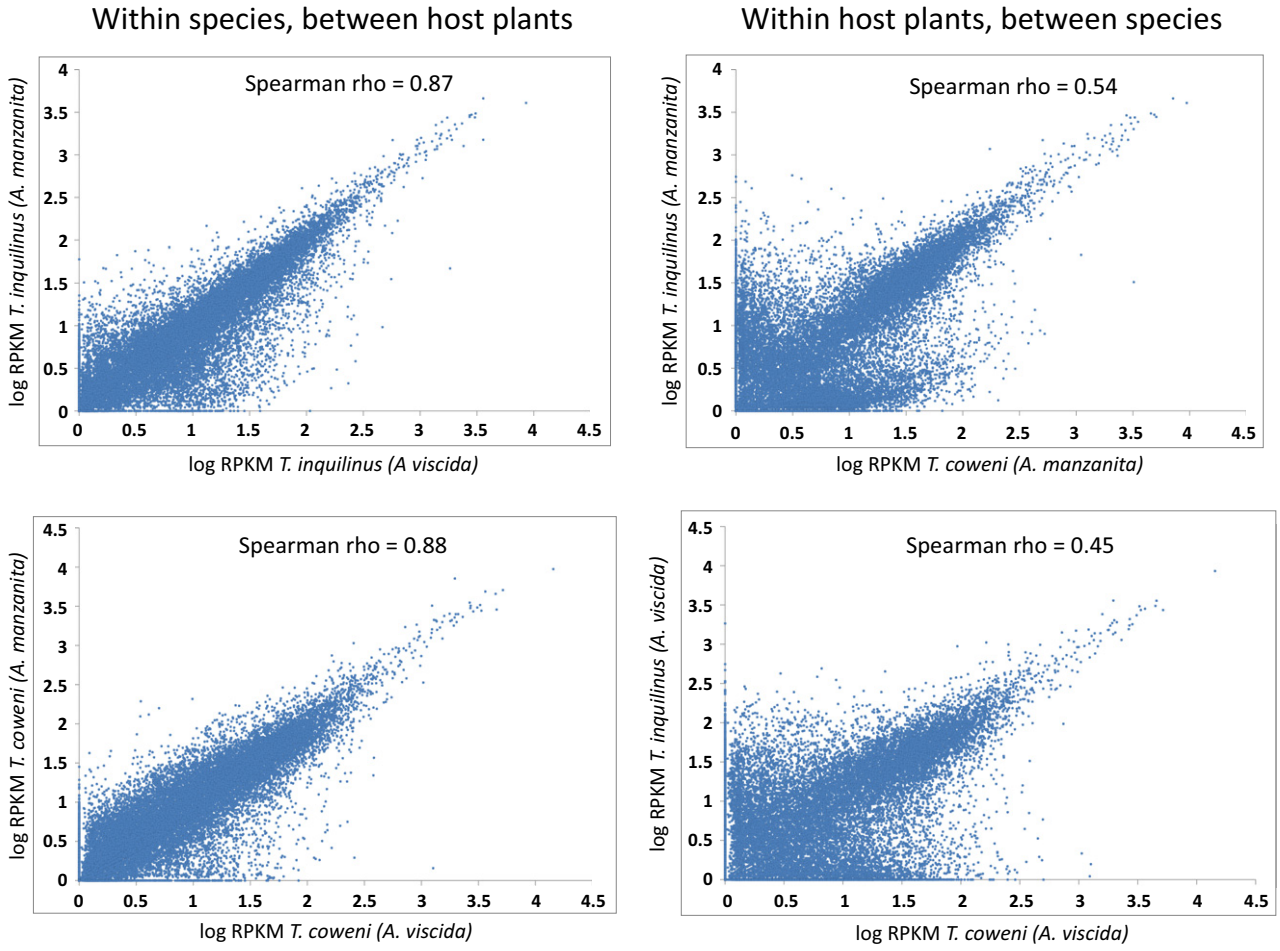

Fig. 4 The correlation within *Tamalia coweni* gene expression across host plants (top) vs. between *T. coweni* and *Tamalia inquilinus* on a common host plant (*Arctostaphylos viscida*). Darker shading indicates greater transcript density.

that involve specialized interactions across trophic levels (Price 2005). We used multiple nuclear genes and one endosymbiont marker from *Tamalia* aphids to investigate the evolution of an unusual within-guild interaction, that is, inquilinism, and to describe patterns of host-plant associations and diversification in these specialist herbivores. The inquiline, *Tamalia inquilinus*, invades and co-occupies galls of close relatives on its host plants, but does not induce galls itself. Gall-forming *Tamalia* have radiated on manzanita and related host plants, and one scenario for the origin and diversification of the inquiline is that it diverged in sympatry with its gall-forming relative. We identified a single origin of inquilines on *Arctostaphylos*, and although we did not find unambiguous evidence of the order of colonization of *Tamalia* host plants, our results generally support the proposition by Miller & Crespi (2003) that the parasitic origin of *T. inquilinus* was associated with a host-plant shift in gall-inducing lineages from the genera *Arbutus* and *Comarostaphylis* to *Arctostaphylos*. We also found that the inquiline exhibits a stronger pattern of divergence across host plants, perhaps reflecting its parasitic life history and the effects of founder events and small effective population sizes on genetic variation (Huyse *et al.* 2005). Our results suggest that specialized, within-guild interactions can thus promote phylogenetically synchronous patterns of diversification in communities of plant-feeding insects resembling the phylogenetic cascades of some specialized trophic interactions between phytophagous insects and their predators and parasites (Forister & Feldman 2010).

The radiation of Tamalia on its host plants

The taxonomic relationships of early-diverging lineages of aphids such as *Tamalia* remain poorly resolved. *Tamalia* itself harbours abundant diversity on Mexican and Central American *Arbutus* and *Comarostaphylis* not included in this study. With these caveats, it appears that the radiation of *Tamalia* is generally consistent with

patterns of radiation in its host plants (Hileman *et al.* 2001; Wahlert *et al.* 2009). *Tamalia* likely originated on *Arbutus arizonica*, which places the gall inducers on *Comarostaphylis diversifolia* sister to the clade on *Arctostaphylos* spp., consistent with earlier results based on mtDNA (Miller & Crespi 2003). Extant *Arctostaphylos* species are of recent origin, likely originating in the late Pliocene (Wahlert *et al.* 2009). The differentiation of *Tamalia* on *Arctostaphylos* has thus been relatively rapid, as is the case in other gall-forming insect groups (Price 2005). Three previously identified taxa (*Tamalia* Species A, *Tamalia* Species B and *Tamalia* Species C) are substantiated in our analysis, while one described species (*T. dicksoni*) is not (Remaudière & Stroyan 1984). *Tamalia dicksoni* may be synonymous with *Tamalia coweni*, although further work will be needed to determine this. *Tamalia dicksoni* has been described only from *Arctostaphylos glandulosa* and *Arctostaphylos pringlei* in California. Both may represent recent colonization events, and incomplete lineage sorting may partially account for the lack of resolution in our study.

Host-associated diversification in Tamalia gall inducers and inquilines

Under the hypothesis of sequential radiation between plants, herbivores and their natural enemies (Abrahamson & Blair 2008), divergence of the host-plant lineage fuels the divergence of the specialist herbivores. This, in turn, fuels ecological radiation of parasites that specialize on herbivores (Forbes *et al.* 2009; Feder & Forbes 2010). Because gall induction is an intricate process (Wool 2004), host-associated diversification (HAD) is probably the rule among lineages of gall inducers. An unusual pattern revealed by our study is seen in the tendency of inquilines to be *more* strongly associated with host-plant species than are gall inducers: this is evident in the phylogenies, as well as in the haplotype networks, and AMOVA and Mantel analyses. Galling aphids have evolved specialized feeding functions requiring an intimate biochemical association between herbivore and plant (Floate & Whitham 1993), and we expected HAD to be greater for gall-forming *Tamalia* than for *T. inquilinus*, which does not express traits for gall induction. However, our data suggest the opposite pattern: *T. inquilinus* expresses more, not less, divergence across host plants. Geographic isolation probably accounts for some of the differentiation we found across host plants. However, gall-forming *Tamalia* likely experience geographic isolation as well, and with regard to geographic subdivision, it is not immediately obvious why *T. inquilinus* would experience greater subdivision than gall formers. When we directly compared host and inquiline samples across three host plants sampled from identical populations (*A. viscida*, *A. manzanita*, and *A. patula*), the pattern of relatively greater differentiation in *T. inquilinus* was still substantial, indicating that either HAD is greater in the inquiline than in its hosts, or perhaps a set of population-genetic processes are operating in *T. inquilinus* that differs from that of its hosts. Partial Mantel tests indicate that, relative to host plants, population structure explains relatively little of the genetic variation in *Tamalia*. Hence, for reasons that are not entirely clear, HAD is greater in the parasitic lineage than in its host, and intraspecific patterns of differentiation appear to be occurring relatively rapidly. This is in contrast to the pattern seen in some brood-parasitic birds and cleptoparasitic bees, in which parasitic lineages experience greater rates of extinction and reduced rates of diversification, presumably because host specificity imposes greater risk upon the parasites than on their ecologically more labile hosts (Dunn *et al.* 2009; Krüger *et al.* 2009; Litman *et al.* 2013). Perhaps the assumption that *T. inquilinus* is relatively free from constraints imposed by host plants is incorrect. *Tamalia dicksoni*, for example, induces galls on *A. glandulosa* and *A. pringlei* in California and is parasitized by inquilines. Both host-plant species appear well defended, with abundant glandular trichomes on the foliage, potentially reducing their suitability as hosts to non-gall-forming insects such as *T. inquilinus* (Andres & Connor 2003). Alternatively, there may be demographic aspects of the parasitic life history that contribute to greater population subdivision than that seen in the gall-forming host aphids. Limited dispersal, smaller effective population sizes, or both, coupled with extinction rates similar to those of other *Tamalia* aphids, may be driving HAD in inquiline aphids, rather than stronger selective constraints inherent in the host plants themselves. Ongoing field studies (D. G. Miller, unpublished data) are providing evidence, first, that inquilines undergo more localized dispersal and, second, that inquiline populations experience greater fluctuations than do gall inducers, both phenomena consistent with a genetic bottleneck hypothesis (Wright 1931; Akimoto 1988).

Potential origins of inquilinism

To date, *Tamalia* is the sole known aphid genus comprising obligate, congeneric inquiline species. General questions regarding the *Tamalia* inquilines include how they arose, how specialized they are and how host-plant associations and host shifts have constrained or fostered their divergence. Emery's Rule holds that social parasites evolved from their hosts, and has evidently operated in various taxonomic groups, including *Myrmica* ants (Savolainen & Vepsäläinen 2003), allodapine and stingless bees (Smith *et al.* 2007; Quezada-Euán

et al. 2013), curculionid beetles (Jordal *et al.* 2000), and even primates (Jones 2005). Miller & Crespi (2003) hypothesized a single, sympatric origin of inquilines within the *Tamalia* lineage on *Arctostaphylos*, in accord with Emery's Rule. Our results support their scenario. However, as with most cases that purportedly support Emery's Rule, the mechanisms underlying divergent selection on host and inquiline populations are unknown. For *Tamalia*, we can outline a model in which, in conjunction with divergent selection imposed by host plants, the tendency of *Tamalia* aphids to communally share galls has acted as a pre-adaptation for inquilinism (Miller 2005). Intraspecific, facultative inquilinism may have facilitated the interspecific, obligate inquiline habit, as illustrated in *Myrmica* ants (Savolainen & Vepsäläinen 2003). Intraspecific inquilinism occasionally occurs in *Tamalia* when multiple foundresses representing different instars cohabit galls: evidently, older females initiate galls, after which younger foundresses join them in the mature gall (D. G. Miller, unpublished data). This observation is consistent with a hypothesis put forward by Hamilton (1987) that interspecific inquilinism originates in gall-forming aphids via the tendency of juvenile aphids (i.e. first instars) to have more generalized, flexible morphologies than adults for moving between and colonizing galls. If so, one mechanism by which reproductive isolation may have emerged between *T. inquilinus* and its hosts is the phenological displacement occurring between adults that formed galls and juveniles that tended to move between and colonize them subsequently. Manzanitas are adapted to the chaparral and coastal regions of Western North America, and once their waxy leaves mature, *Tamalia* aphids cannot initiate galls (Sholes & Beatty 1987). Gall-forming *Tamalia* are thus tightly constrained by the phenology of their host plants, whereas inquiline life cycles are constrained by narrow windows of opportunity to enter galls themselves, occurring later in the season (Miller 2005). The sexual adult males and females of *T. coweni* are larviposited by the gall foundresses and depart the gall upon adulthood. *Tamalia inquilinus* can remain within the gall long after the sexuals of the gall foundress depart, and produce its winged sexuals later (Miller & Crespi 2003). These discrete phenological differences, coupled with host-associated differentiation, may be the key mechanisms by which divergent selection drives reproductive isolation in inquilines (Miller & Crespi 2003).

If inquilines do express evidence of specificity to their gall-forming hosts distributed across populations and host plants, are inquilines tracking their host plants, or the host aphids? It is unlikely that the driver of specialization is the aphids themselves. It is more likely that as yet undescribed, aspects of adaptation to host plants drive ecological divergence in the inquilines at least as much as they do in the gall formers (Döring 2014). Although speculative, the RNA-seq analysis provides some hints of how inquilines respond to host plants and suggests that assumptions that inquilines experience fewer host-plant constraints than do gall formers are misplaced. Moran & Jarvik (2010) recently described the presence in aphids of the structural genes needed for carotenoid biosynthesis, the metabolites of which can act to quench singlet oxygen species produced by plants during aphid feeding (Giordanengo *et al.* 2010). Carotenoids are pigments and consist of long, conjugated chains of double bonds that can be biochemically modified and cleaved in various ways, conferring different activities (Cazzonelli 2011; Walter & Strack 2011). Genes involved in carotenoid biosynthesis were overexpressed in one of the inquiline samples, constituting one of the most highly enriched genes in our data set. Overexpression of carotenoids in inquilines reflects a difference in how inquilines and gall formers induce and mitigate plant defences, and a potential indirect, costly effect of their cohabitation within the galls of their hosts (Fretwell 1978; Carroll & Berenbaum 2006). As well, there were the differences in the expression of genes involved in metabolism and reproduction, possibly indicating life history differences between the two species. The upregulation of genes involved in metabolism and reproduction in the inquilines may reflect their extended period of growth and reproduction on their host plants, relative to their host aphids, and represent transcriptional correlates of the phenological differences between inquilines and gall formers described above. Reciprocal transplant trials of inquilines between host-plant species, coupled with experiments to profile transcriptional responses, may reveal subtle adaptations that help shed light on these questions (Akimotos 1990).

Conclusions

Tamalia gall formers and their inquilines are radiating across North American *Arctostaphylos* in a nearly synchronous fashion, constituting an unusual example in which the diversity of host species is indirectly coupled to that of their parasites in the absence of direct, consumptive interactions. The circumstances promoting a greater degree of host-associated differentiation in *Tamalia* social parasites than in their gall-inducing hosts are unclear, but the loss of gall-inducing traits at the origins of inquilinism appears to be likely associated with a host-plant shift to *Arctostaphylos* from the allied genera *Arbutus* or *Comarostaphylis* in the *Tamalia* lineage. This unique network of host plant, herbivore and social parasite lineages is an example of how specialized inter-

actions with plants can promote patterns of linked, within-guild diversification in the parasites dependent upon them.

Acknowledgements

This work was supported in part through a Research Opportunity Award supplement to a National Science Foundation award (DEB-1029374) to DGM and PA, and NSF award IOS-1147033 to PA. Lodging and access to populations of insects and host plants was provided by the University of California Santa Cruz Island Reserve and at the Sweeney Granite Mountains Desert Research Center. Additional access to study samples was granted by the Big Chico Creek Ecological Reserve, Stunt Ranch Santa Monica Mountains Reserve and the Mojave National Preserve. We thank Tom Parker and Kristina Schierenbeck for critical reviews of the manuscript.

References

Abrahamson WG, Blair CP (2008) Sequential radiation through host-race formation: herbivore diversity leads to diversity in natural enemies. In: *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects* (ed. Tilmon KJ), pp. 188–202. University of California Press, Berkeley, California.

Abrahamson WG, Blair CP, Eubanks MD, Morehead SA (2003) Sequential radiation of unrelated organisms: the gall fly *Eurosta solidaginis* and the tumbling flower beetle *Mordellistena convicta*. *Journal of Evolutionary Biology*, **16**, 781–789.

Abrams PA (2000) Character shifts of prey species that share predators. *The American Naturalist*, **156**, S45–S61.

Akimoto S (1988) The evolution of gall parasitism accompanied by a host shift in the gall aphid, *Eriosoma yangi* (Homoptera: Aphidoidea). *Biological Journal of the Linnean Society*, **35**, 297–312.

Akimoto S (1989) Gall-invading behavior of *Eriosoma* aphids (Homoptera, Pemphigidae) and its significance. *Japanese Journal of Entomology*, **57**, 210–220.

Akimotos S (1990) Local adaptation and host race formation of a gall-forming aphid in relation to environmental heterogeneity. *Oecologia*, **83**, 162–170.

Althoff DM, Segraves KA, Johnson MTJ (2014) Testing for coevolutionary diversification: linking pattern with process. *Trends in Ecology & Evolution*, **29**, 82–89.

Altschul SF, Madden TL, Schäffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Andres MR, Connor EF (2003) The community-wide and guild-specific effects of pubescence on the folivorous insects of manzanitas *Arctostaphylos* spp. *Ecological Entomology*, **28**, 383–396.

Bailey JK, Whitham TG (2003) Interactions among elk, aspen, galling sawflies and insectivorous birds. *Oikos*, **101**, 127–134.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Blackman RL, Eastop VF (2006) *Aphids on the World's Herbaceous Plants and Shrubs*. Wiley, Chichester, UK.

Buschinger A (1986) Evolution of social parasitism in ants. *Trends in Ecology and Evolution*, **1**, 155–160.

Carroll MJ, Berenbaum MR (2006) Lutein sequestration and furanocoumarin metabolism in parsnip webworms under different ultraviolet light regimes in the montane west. *Journal of Chemical Ecology*, **32**, 277–305.

Cazzonelli CI (2011) Goldacre review: carotenoids in nature: insights from plants and beyond. *Functional Plant Biology*, **38**, 833–847.

Chapman TW, Crespi BJ, Perry SP (2008) The evolutionary ecology of eusociality in Australian gall thrips: a "model clades" approach. In: *Ecology of Social Evolution* (eds Korb J, Heinze J), pp. 57–83. Springer, Berlin, Heidelberg.

Clark MA, Moran NA, Baumann P, Wernegreen J (2000) Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution*, **54**, 517–525.

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

Craig TP, Itami JK, Horner JD (2007) Geographic variation in the evolution and coevolution of a tritrophic interaction. *Evolution*, **61**, 1137–1152.

Davies NB, Brooke MD (1988) Cuckoos versus reed warblers: adaptations and counter adaptations. *Animal Behaviour*, **36**, 262–284.

Dieckmann U, Doebeli M, Metz JAJ, Tautz D (2012) *Adaptive Speciation*. Cambridge University Press, Cambridge, UK.

Döring TF (2014) How aphids find their host plants, and how they don't. *The Annals of Applied Biology*, **165**, 3–26.

Drummond CS, Xue H-J, Yoder JB, Pellmyr O (2009) Host-associated divergence and incipient speciation in the yucca moth *Prodoxus coloradensis* (Lepidoptera: Prodoxidae) on three species of host plants. *Heredity*, **105**, 183–196.

Dunn RR, Harris NC, Colwell RK, Koh LP, Sodhi NS (2009) The sixth mass coextinction: are most endangered species parasites and mutualists? *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 3037–3045.

Emerson BC, Kolm N (2005) Species diversity can drive speciation. *Nature*, **434**, 1015–1017.

Emery C (1909) Über den Ursprung der dulotischen, parasitischen und myrmekophilen Ameisen. *Biologisches Centralblatt*, **29**, 352–362.

Ersts PJ (2015) *[Internet] Geographic Distance Matrix Generator (version 1.2.3)*. American Museum of Natural History, Center for Biodiversity and Conservation. Available from http://biodiversityinformatics.amnh.org/open_source/gdmg.

Evans LM, Allan GM, Meneses N, Max TL, Whitham TG (2013) Herbivore host-associated genetic differentiation depends on the scale of plant genetic variation examined. *Evolutionary Ecology*, **27**, 65–81.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Farrell BD, Mitter C, Futuyma DJ (1992) Diversification at the insect-plant interface: insights from phylogenetics. *BioScience*, **42**, 34–42.

Feder JL, Forbes AA (2010) Sequential speciation and the diversity of parasitic insects. *Ecological Entomology*, **35**, 67–76.

Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, Zhang Y (2012) GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics*, **28**, 2782–2788.

Floate KD, Whitham TG (1993) The "hybrid bridge" hypothesis: host shifting via plant hybrid swarms. *The American Naturalist*, **141**, 651–662.

Forbes AA, Powell TH, Stelinski LL, Smith JJ, Feder JL (2009) Sequential sympatric speciation across trophic levels. *Science*, **323**, 776–779.

Forister ML, Feldman CR (2010) Phylogenetic cascades and the origins of tropical diversity. *Biotropica*, **43**, 270–278.

Foster WA (2002) Soldier aphids go cuckoo. *Trends in Ecology and Evolution*, **17**, 199–200.

Fretwell J (1978) The distribution of carotenoids in insects. In: *Biochemical Aspects of Plant and Animal Coevolution* (ed. Harborne JB), pp. 277–287. Academic Press, London.

Gilbert JDJ, Mound LA, Simpson SA (2012) Biology of a new species of socially parasitic thrips (Thysanoptera: Phlaeothripidae) inside *Dunatothrips* nests, with evolutionary implications for inquilinism in thrips. *Biological Journal of the Linnean Society*, **107**, 112–122.

Giordanengo P, Brunissen L, Rusterucci C *et al.* (2010) Compatible plant-aphid interactions: how aphids manipulate plant responses. *Comptes Rendus Biologies*, **333**, 516–523.

Hamilton WD (1987) Kinship, recognition, disease and intelligence: constraints of social evolution. In: *Animal Societies: Theories and Facts* (eds Itô Y, Brown JL, Kikkawa J), pp. 81–102. Japan Scientific Society Press, Tokyo, Japan.

Hembry DH, Yoder JB, Goodman KR (2014) Coevolution and the diversification of life. *The American Naturalist*, **184**, 425–438.

Hileman LC, Vasey MC, Parker VT (2001) Phylogeny and biogeography of the Arbutoideae (Ericaceae): implications for the Madrean-Tethyan hypothesis. *Systematic Botany*, **26**, 131–143.

Holt RD, Lawton JH (1994) The ecological consequences of shared natural enemies. *Annual Review of Ecology, Evolution, and Systematics*, **25**, 495–520.

Huang MH, Dornhaus A (2008) A meta-analysis of ant social parasitism: host characteristics of different parasitism types and a test of Emery's rule. *Ecological Entomology*, **33**, 589–596.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Huyse T, Poulin R, Théron A (2005) Speciation in parasites: a population genetics approach. *Trends in Parasitology*, **21**, 469–475.

Janz N, Nylin S, Wahlberg N (2006) Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evolutionary Biology*, **6**, 4.

Jones CB (2005) Social parasitism in mammals with particular reference to neotropical primates. *Mastozoología Neotropical*, **12**, 19–35.

Jordal BH, Normark BB, Farrell BD (2000) Evolutionary radiation of an inbreeding haplodiploid beetle lineage (Curculionidae, Scolytinae). *Biological Journal of the Linnean Society*, **71**, 483–499.

Karvonen A, Seehausen O (2012) The role of parasitism in adaptive radiations—when might parasites promote and when might they constrain ecological speciation? *International Journal of Ecology*, **2012**, Article ID: 280169.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059–3066.

Krüger O, Sorenson MD, Davies NB (2009) Does coevolution promote species richness in parasitic cuckoos? *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 3871–3879.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Langerhans RB (2007) Evolutionary consequences of predation: avoidance, escape, reproduction, and diversification. In: *Predation in Organisms: A Distinct Phenomenon* (ed. Elewa AMT), pp. 177–220. Springer-Verlag, Heidelberg.

Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Litman JR, Praz CJ, Danforth BN, Griswold TL, Cardinal S (2013) Origins, evolution, and diversification of cleptoparasitic lineages in long-tongued bees. *Evolution*, **67**, 2982–2998.

Lohse M, Bolger AM, Nagel A *et al.* (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, **40**, W622–W627.

Lowe RM, Ward SA, Crozier RH (2002) The evolution of parasites from their hosts: intra- and interspecific parasitism and Emery's rule. *Proceedings of the Royal Society of London B: Biological Sciences*, **269**, 1301–1305.

MacArthur RH (1972) *Geographical Ecology*. Harper and Row, New York, New York.

Maddison WP, Maddison DR (2015) *Mesquite: a modular system for evolutionary analysis*. *Version 3.03*. Available from http://mesquiteproject.org.

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **16**, 3448–3449.

Miller DG III (1998a) Life history, ecology, and communal gall occupation in the manzanita leaf-gall aphid, *Tamalia coweni* (Cockerell) (Homoptera: Aphididae). *Journal of Natural History*, **32**, 351–366.

Miller DG III (1998b) Consequences of communal gall induction and a test for kin discrimination in the aphid *Tamalia coweni* (Cockerell) (Homoptera: Aphididae). *Behavioral Ecology and Sociobiology*, **43**, 95–103.

Miller DG III (2004) The ecology of inquilinism in communally parasitic *Tamalia* aphids (Hemiptera: Aphididae). *Annals of the Entomological Society of America*, **97**, 1233–1241.

Miller DG III (2005) Ecology and radiation of galling aphids (*Tamalia*; Hemiptera: Aphididae) on their host plants (Ericaceae). *Basic and Applied Ecology*, **6**, 463–469.

Miller DG III, Crespi B (2003) The evolution of inquilinism, host-plant use, and mitochondrial substitution rates in *Tamalia* gall aphids. *Journal of Evolutionary Biology*, **16**, 731–743.

Miller DG III, Sharkey MJ (2000) An inquiline species of *Tamalia* co-occurring with *Tamalia coweni* (Homoptera: Aphididae). *Pan Pacific Entomologist*, **76**, 77–86.

Moran NA, Jarvik T (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, **328**, 624–627.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford.

Nováková E, Hypša V, Klein J, Foottit RG, von Dohlen CD, Moran NA (2013) Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Molecular Phylogenetics and Evolution*, **68**, 42–54.

Oliveros JC (2007–2015) *Venny. An interactive tool for comparing lists with Venn's diagrams*. Available from http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Price PW (2005) Adaptive radiation of gall-inducing insects. *Basic and Applied Ecology*, **6**, 413–421.

Quezada-Euán JJG, Ramírez J, Eltz T, Pokorny T, Medina R, Monsreal R (2013) Does sensory deception matter in eusocial obligate food robber systems? A study of *Lestrimelitta* and stingless bee hosts. *Animal Behavior*, **85**, 817–823.

Remaudière G, Stroyan HLG (1984) Un *Tamalia* nouveau de Californie (USA). Discussion sur les *Tamaliinae* subfam. nov. (Hom. Aphididae). *Annales de la Société entomologique de France (N.S.)*, **20**, 93–103.

Ronquist F (1994) Evolution of parasitism among closely related species: phylogenetic relationships and the origin of inquilinism in gall wasps (Hymenoptera, Cynipidae). *Evolution*, **48**, 241–266.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed model. *Bioinformatics*, **19**, 1572–1574.

Roof JB (1967) *Arctostaphylos mewukka*, a hybrid. *Four Seasons*, **2**, 16.

Savolainen R, Vepsäläinen K (2003) Sympatric speciation through intraspecific social parasitism. *Proceedings of the National Academy of Sciences of the USA*, **100**, 7169–7174.

Schierenbeck KA, Stebbins GL, Patterson RW (1992) Morphological and cytological evidence for polyphyletic allopolyploidy in *Arctostaphylos mewukka* (Ericaceae). *Plant Systematics and Evolution*, **179**, 187–205.

Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, New York, New York.

Shannon P, Markiel A, Ozier O *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **11**, 2498–2504.

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, **51**, 492–508.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Sholes ODV, Beatty SW (1987) Influence of host phenology and vegetation on the abundance of *Tamalia coweni* galls (Homoptera: Aphididae) on *Arctostaphylos insularis* (Ericaceae). *American Midland Naturalist*, **118**, 198–204.

Silvestro M (2012) raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution*, **12**, 335–337.

Smith JA, Tierney SM, Park YC, Fuller S, Schwarz MP (2007) Origins of social parasitism: the importance of divergence ages in phylogenetic studies. *Molecular Phylogenetics and Evolution*, **43**, 1131–1137.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

Stamatakis A (2006) RAxML-VI-HPC: maximum-likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Stireman JO III, Nason JD, Heard S (2005) Host-associated genetic differentiation in phytophagous insects: general phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. *Evolution*, **59**, 2573–2587.

Stireman JO, Devlin H, Carr TG, Abbot P (2010) Evolutionary diversification of the gall midge genus *Asteromyia* (Cecidomyiidae) in a multitrophic ecological context. *Molecular Phylogenetics and Evolution*, **54**, 194–210.

Stoddard MC, Kilner RM (2013) The past, present and future of 'cuckoos versus reed warblers'. *Animal Behaviour*, **85**, 693–699.

Strong DR, Lawton JH, Southwood R (1984) *Insects on Plants: Community Patterns and Mechanisms*. Harvard University Press, Cambridge, Massachusetts.

Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*. Sinauer Associates, Sunderland, Massachusetts.

Taylor BG, Miller DG (2014) High mean relatedness among communally-galling *Tamalia* aphids revealed by AFLP analysis. *Insectes Sociaux*, **61**, 395–402.

Violle C, Enquist BJ, McGill BJ *et al.* (2012) The return of the variance: intraspecific variability in community ecology. *Trends in Ecology & Evolution*, **27**, 245–253.

Wahlert GA, Parker VT, Vasey MC (2009) A phylogeny of *Arctostaphylos* (Ericaceae) inferred from nuclear ribosomal ITS sequences. *Journal of the Botanical Research Institute of Texas*, **3**, 673–682.

Walter MH, Strack D (2011) Carotenoids and their cleavage products: biosynthesis and functions. *Natural Product Reports*, **28**, 663–692.

Wool D (2004) Galling aphids: specialization, biological complexity, and variation. *Annual Review of Entomology*, **49**, 175–192.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Xie Y, Wu G, Tang J *et al.* (2013) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *ArXiv*, 1305.6760.

Zink AG (2000) The evolution of intraspecific brood parasitism in birds and insects. *The American Naturalist*, **155**, 395–405.

P.A. and D.G.M. conceived and designed the study; D.G.M. collected and prepared the insect samples; P.A., S.P.L. and H.E. performed the molecular laboratory work; P.A. and D.R. performed the RNA-seq analyses; P.A. and S.P.L. performed the phylogenetic analyses; P.A. and D.G.M. wrote the manuscript.

Data accessibility

DNA sequences: GenBank accessions KM604232–KM604313; KM604314–KM604395; KM604396–KM604477; KM604478–KM604559; KM604560–KM604641; NCBI SRA: SRX1305377; SRX1305445; SRX1305446; SRX1305421; SRX1305282; SRX1304838.

Alignments, trees, sequence assemblies, annotations and GFOLD summaries: Dryad doi: 10.5061/dryad.3ck0v.

Supporting information

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

Fig. S1 Maximum-likelihood consensus tree from *Buchnera* groEL, rooted with three species from a closely related subfamily (Thelaxinae).

Fig. S2 Bayesian consensus tree from *Buchnera* groEL, rooted with three species from a closely related subfamily (Thelaxinae).

Fig. S3 The unrooted majority-rule consensus from the maximum-likelihood analysis of the concatenated nuclear gene matrix.

Fig. S4 The unrooted phylogram of the best tree from the maximum-likelihood analysis of PEPCK.

Fig. S5 Haplotype networks for *Tamalia coweni* and *Tamalia inquilinus*, derived from *Buchnera* groEL.

Fig. S6 Patterns of nucleotide variation among (dark shading) and between (light shading) host plants of *Tamalia coweni* gall-inducers and *Tamalia inquilinus* samples.

Fig. S7 Patterns of nucleotide variation among (dark shading) and between (light shading) populations of *Tamalia coweni* gall-inducers and *Tamalia inquilinus* samples sampled from common sites across three host plants.

Table S1 Sequenced *Tamalia* samples and collection information used in all phylogenetic and population genetic analyses, except PEPCK intron.

Table S2 Sequenced *Tamalia* samples used in the analysis of the PEPCK intron.

Table S3 Primer sequence and associated information for genes used in this study.

Table S4 Models and partitions for Maximum Likelihood and Bayesian analyses conducted on the concatenated matrix, groEL, and the PEPCK intron.

Table S5 Summary statistics for the assembly of RNA sequencing for all scaffolds and for unique scaffolds only.

Table S6 The top 100 genes with functional annotations that were significantly differentially expressed between *Tamalia coweni* and *Tamalia inquilinus* on *Arctostaphylos viscida*.

Table S7 The top 100 genes with functional annotations that were significantly differentially expressed between *Tamalia coweni* and *Tamalia inquilinus* on *Arctostaphylos manzanita*.

Table S8 Sequence polymorphism statistics for Plc21c, PEPCK and groEL from *Tamalia coweni* and *Tamalia inquilinus*.

Table S9 AMOVA table for *Tamalia coweni* and *Tamalia inquilinus* collected from a common host plant species (*Arctostaphylos viscida*) sampled from three different populations.