Phylogenetic congruence of mealybugs and their primary endosymbionts

D. A. DOWNIE* & P. J. GULLAN†

*Department of Zoology and Entomology, Rhodes University, Grahamstown, South Africa
†Department of Entomology, University of California, Davis, CA, USA

Abstract

Tight interactions between unrelated organisms such as is seen in plant–insect, host–parasite, or host–symbiont associations may lead to speciation of the smaller partners when their hosts speciate. Totally congruent phylogenies of interacting taxa have not been observed often but a number of studies have provided evidence that various hemipteran insect taxa and their primary bacterial endosymbionts share phylogenetic histories. Like other hemipterans, mealybugs (Pseudococcidae) harbour multiple intracellular bacterial symbionts, which are thought to be strictly vertically inherited, implying codivergence of hosts and symbionts. Here, robust estimates of phylogeny were generated from four fragments of three nuclear genes for mealybugs of the subfamily Pseudococcinae, and a substantial fragment of the 16S–23S rDNA of their P-endosymbionts. Phylogenetic congruence was highly significant, with 75% of nodes on the two trees identical, and significant correlation of branch lengths indicated coincident timing of cladogenesis. It is suggested that the low level of observed incongruence was influenced by uncertainty in phylogenetic estimation, but evolutionary outcomes other than congruence, including host shifts, could not be rejected.

Introduction

Coevolutionary interactions may be diffuse, involving a number of participants, or specific, involving a pair of species exerting reciprocal selection pressures on each other (Thompson, 1994). Specific interactions involving coevolutionary arms races between antagonists (e.g. Walling, 2000; Karban & Agrawal, 2002) and fine-tuning of mutualisms (e.g. Donaldson, 1997, see references in Cook & Rasplus, 2003) have been of most interest to researchers interested in coevolution. Although the adaptive responses of the players in a coevolutionary dance clearly are relevant to understanding organic diversity, how, and if, the interactants diversify in tandem is equally relevant. Partners in highly specific interactions, such as those between specialized insect herbivores and their host plants, parasites and their hosts, or obligately

mutualistic species, might be expected to diversify in tandem because of coadaptation, or mutual isolation from close relatives. The form of divergence may be quite different in these cases, as some interactions should lead to adaptive divergence between sister taxa (insects on plants), whereas others may lead to divergence simply as a consequence of a highly specialized partner being 'captured' by its host and isolated from its sister taxon (some parasites). Nonadaptive divergence by parasite capture is contrary to certain views of the speciation process (Schluter, 2001; Wu, 2001) but may be relatively common and is consistent with the classic allopatric scenario (Dobzhansky, 1937). Thus additional examples of codivergence/cospeciation (or its absence) will lead to a more general picture of organic diversification.

The evidence for cospeciation of interacting partners has been scant in spite of the existence of a wealth of tight interactions. In most cases the linkage between host and associate phylogenies has been loose, exhibiting a mix of co-speciation and other outcomes (Page, 2002). Strict cospeciation is seldom observed (Johnson *et al.* 2003). The best evidence for cospeciation in interacting

Correspondence: D. A. Downie, Department of Zoology and Entomology, Rhodes University, Grahamstown 6140, South Africa.
Tel.: 027 046 603 8086; fax: 027 046 622 8959;
e-mail: d.downie@ru.ac.za

partners has come from mammals and their parasites (Hafner & Page, 1995), certain plants and their pollinating seed-predatory insects (e.g. figs and fig wasps, Cook & Rasplus, 2003) and herbivorous insects in the order Hemiptera, suborder Sternorrhyncha, and their primary endosymbiotic bacteria (P-endosymbionts). Recent phylogenetic studies in aphids (Munson *et al.*, 1991; Clark *et al.*, 2000), psyllids (Thao *et al.*, 2000; Spaulding & von Dohlen, 2001), whiteflies (Clark *et al.*, 1992; Thao & Baumann, 2004), and leafhoppers (Moran *et al.*, 2003) have shown close congruence between host and endosymbiont phylogenies, and between different symbiont genomes (Funk *et al.*, 2000; Thao *et al.*, 2002).

It should be noted that most studies cited above have focused on a higher taxonomic level and have not truly validated the claimed strict cospeciation. The sole study that focused on the level of species within a genus and would thus be able to test for strict cospeciation failed to find it, although the authors suggested that incongruence was more apparent than real because of limitations of the statistical tests (Clark *et al.*, 2000). Clark *et al.* offered the useful insight that lack of resolution and conflict among data sets could obscure a true pattern of strict phylogenetic congruence.

Congruence of phylogenetic history is expected in these hemipteran–endosymbiont systems because of the vertical inheritance of the endosymbiont partners that are housed within specialized polyploid cells (bacteriocytes) in organs (bacteriomes) closely associated with the insect midgut. Typically, during oogenesis a small group of bacterial cells migrates to the ovarioles, passing into the oocyte, and thence to the developing embryo (Buchner, 1965; Moran & Telang, 1998). No other form of transmission has been documented. The P-endosymbionts supplement the nutrient poor diet of Hemiptera that feed on plant sap by synthesizing critical amino acids. Most plant-feeding Hemiptera also house secondary, maternally transmitted bacterial endosymbionts, which are usually not associated with bacteriomes. Their locations within their insect hosts are variable, their function uncertain, and their phylogenies largely incongruent with that of their hosts, suggesting frequent horizontal transmission (Russell *et al.*, 2003).

A new level of complexity in hemipteran–endosymbiont interactions was revealed by research showing that the P-endosymbionts of mealybugs (Pseudococcidae), designated as *Candidatus Tremblaya princeps* (Thao *et al.*, 2002), housed cells of the secondary symbionts within their own cells (von Dohlen *et al.*, 2001), and that the two kinds of endosymbionts shared congruent phylogenetic histories (Thao *et al.*, 2002). A phylogeny for pseudococcid P-endosymbionts using 16S and 23S rDNA sequences (Thao *et al.*, 2002) did not address the phylogeny of the mealybug hosts – only four mealybug species were sequenced for the mitochondrial gene cytochrome oxidase I (COI) as an approximate check of concordance of the mealybug and bacterial relationships.

Here, we explicitly investigate the purported phylogenetic congruence of mealybugs and their P-endosymbionts. DNA sequences were acquired from three nuclear genes for 21 of the 22 mealybug hosts of the P-endosymbionts examined by Thao *et al.* (2002). Using these and the 16S–23S rDNA sequences of the P-endosymbionts already available, robust estimates of phylogeny were generated at the level of host genera within a subfamily (Pseudococcinae, see Downie & Gullan, 2004). We test the null hypothesis of strict phylogenetic congruence of mealybugs and their P-endosymbionts, and consider how any incongruence between mealybug and P-endosymbiont phylogenies might be explained.

Materials and methods

Specimens were obtained preserved in 95–100% ethanol. With the exception of *Vryburgia brevicruris*, mealybugs collected for this study were the same species (in some cases the same samples) used by Thao *et al.* (2002). *Ferrisia gilli* in this study, and in Downie & Gullan (2004), is the same mealybug as the unnamed *Ferrisia* sp. in Thao *et al.* (2002), described by Gullan *et al.* (2003). Collecting details, treatment for identification and vouchering of specimens are described in Thao *et al.* (2002) and Downie & Gullan (2004).

DNA extraction, polymerase chain reaction and sequencing

The DNA sequences were obtained for 1513 to 1522 bp of 16S rDNA and 2527 to 2534 bp of 23S rDNA for the P-endosymbionts of 21 mealybug taxa as described in Thao *et al.* (2002). Protocols for sequencing four fragments of three nuclear genes from the 21 mealybug taxa, 316 bp of the protein coding gene EF-1α, 255 bp of the rDNA 28S D2 expansion region, 712 bp of the 28S D10 expansion region, and 678 bp of the rDNA 18S gene, can be found in Gullan *et al.* (2003) and Downie & Gullan (2004). For *Nipaecoccus exocarpi* sequences were obtained only for the EF-1α gene fragment. *Puto yuccae* (Putoidae) and *Icerya purchasi* (Margarodidae) were sequenced as outgroups for mealybugs and *Acetobacter intermedius* and *Neisseria meningitidis* were used as outgroups for the P-endosymbionts.

Phylogenetic analysis

Phylogenetic signal in the data was assessed by comparing the g_1 statistic proposed by Hillis & Huelsenbeck (1992) with their critical values. The model of DNA substitution of best fit to the data was found using Modeltest ver. 3.06 (Posada & Crandall, 1998). These parameters were then used to generate phylogenetic trees for individual and combined data sets by maximum likelihood (ML). Heuristic searches were conducted for

both mealybug host and P-endosymbiont sequences using 10 random addition sequence replicates. Nonparametric bootstrapping using 100–200 replicates was used to estimate support for nodes. Additional estimates of phylogeny were generated by maximum parsimony (MP) using 100 random sequence addition replicates in heuristic searches followed by 500 bootstrap replicates to assess node support. Phylogenetic analyses were run with PAUP* ver. 4.0b10 (Swofford, 2003). Analyses were run on individual data sets as well as on combined data but, except where noted, all discussion is based on the combined data set. Incongruence Length Difference tests (ILD test)(Farris *et al.*, 1994) using 500 replicates of heuristic search were used to assess agreement among the gene regions sequenced in mealybugs, as well as between all possible combinations of mealybug and P-endosymbiont data sets. It was not used as a basis for decisions on combining data.

Statistical tests of coevolution

TreeMap analysis

TreeMap ver. 1.0 (Page, 1994) was used to assess the degree of congruence between host and P-endosymbiont trees by estimating the number of codivergences, duplications, sorting events and host shifts, testing the departure from a random expectation of the number of codivergences, and examining the correlation between branch lengths of mealybug and symbiont trees. TreeMap's optimality criterion of maximizing the number of codivergences makes it prone to find phylogenetic congruence in preference to other outcomes. Besides the well-understood outcomes of codivergence and host shifts, duplications result from speciation in one partner but not the in other, and sorting events account for extinctions, or unsampled taxa, along symbiont or parasite phylogenies. Trees were based on the ML estimates for both mealybugs and P-endosymbionts, and, to account for slight differences from ML trees, analyses were run also using all MP trees for both partners, for combined data sets. TreeMap analyses included outgroups but these were pruned from both host and symbiont trees in the figures because of the long branch subtending the bacterial outgroups (the free-living bacteria and coccoids used as outgroups were not associates in any case). An exhaustive search was used to find the best reconstruction, and 1000 replicates of randomization of P-endosymbiont trees, using the proportional to distinguishable model, were run to test the hypothesis that the observed number of codivergences differed significantly from the random expectation. In addition, a pattern of codivergence should lead to a correlation between the branch lengths of host and symbiont trees (different rates of evolution and nonhomology of the genes sampled may preclude direct comparison of absolute branch lengths, but a correlation should still exist). The significance of the correlation between the branch lengths of mealybug and P-endosymbiont trees was evaluated by randomization tests.

SOWH test

The above randomization procedure tests whether congruence differs significantly from a random expectation but does not address the perhaps more germane question of whether incongruence is greater than expected by chance. This hypothesis was tested using the SOWH test (Goldman *et al.*, 2000), originally suggested by Swofford *et al.* (1996). The test is appropriate when *a posteriori* comparisons are made, a situation for which the more widely used KH test may not be appropriate (Goldman *et al.*, 2000). Given a constraint tree, the method uses parametric bootstrapping to generate a null distribution of the difference in likelihoods under the model of sequence evolution used to estimate the original ML tree. Here, 100 data sets of the same length as the mealybug data set were simulated with Seq-Gen ver. 1.2.6 (Rambaut & Grassly, 1996) using parameters of sequence evolution estimated from the original mealybug data using the P-endosymbiont ML tree as a constraint tree (null hypothesis). The difference in log likelihoods ($\delta = \ln L_{\text{MLmealybug}} - \ln L_{\text{MLsymbiont}}$) of the observed mealybug and P-endosymbiont trees, given the mealybug data, was compared with the distribution of δ for the unconstrained vs. null hypothesis trees estimated from each simulated data set. An observed difference outside the 95% point of the simulated distribution would indicate that incongruence was greater than expected by chance alone. It should be noted that the SOWH test allows one to reject or accept the hypothesis that tree topologies differ because of sampling error alone, but does not account for systematic biases in the data.

Partitioned Bremer support

The extent and nature of character congruence among data partitions can be examined in a parsimony framework by means of partitioned Bremer support (PBS), a method devised by Baker & DeSalle (1997) to assign to each data partition the contribution to character support for any given node on a combined analysis tree. A set of constraint trees, each differing only by the collapse of a single node, is evaluated for each data partition. The PBS is the difference in the number of steps for a data partition on the most parsimonious tree (MPT) for the combined data and the shortest tree without a given node. The PBS can be positive or negative, a positive value indicating support for a node, a negative value indicating conflict with that node. The sum of the PBS values will equal the total Bremer support for a node. Here, the entire data set of 6077 bases, combining all the mealybug genes and the P-endosymbiont 16S–23S sequences, was analysed under the same heuristic search settings described above, with all characters given equal weight. The PBS for each node was calculated using TreeRot ver. 2.0 (Sorenson, 1999).

Results

Sequence details for P-endosymbiont 16S–23S rDNA are described in Thao *et al.* (2002) and the mealybug sequences are described in detail in Downie & Gullan (2004). Pair-wise distances among P-endosymbionts ranged from 0.002 to 0.089, and from 0.011 to 0.224 among mealybug hosts (combined data, outgroups excluded). The lower rate of divergence among P-endosymbionts is likely due to the conserved nature of the symbiont genes sequenced rather than a reflection of relative evolutionary rate.

Phylogenetic results

Data sets for both mealybugs and P-endosymbionts were informative: there were 329 and 757 parsimony informative characters in the combined mealybug data set and the 16S–23S P-endosymbiont data set, respectively. Tree statistics are given in Table 1. All individual and combined datasets had significant phylogenetic signal (mealybug 18S: $g_1 = -1.020$; 28S D2: $g_1 = -0.861$; 28S D10: $g_1 = -0.697$; EF-1α: $g_1 = -0.603$; combined: $g_1 = -0.554$; P-endosymbiont 16S–23S: $g_1 = -1.454$; all $P < 0.01$).

Mealybugs
Significant incongruence was found among the four data partitions for the mealybugs (ILD test, $P < 0.01$). Pairwise comparisons showed that EF-1α and 28S D2 were the cause of incongruence ($P < 0.05$), differing in how the two *Planococcus* species were placed (sister to the *Ferrisia* group (Downie & Gullan, 2004) for 28S D2, sister to all other mealybugs excepting the two *Maconellicoccus* species for EF-1α. This comparison was not significant after a Bonferroni correction however, and no other comparisons were significant. The only disagreements found among data partitions at supported nodes were sister relationships of *Paracoccus nothofagicola* and *Australicoccus grevilleae* for the 18S fragment and *Melanococcus albizziae* and *Amonostherium lichtensioides* for the 28S D10 fragment that were contradicted by other partitions.

The model of sequence evolution used for ML analysis of the combined dataset was the Tamura Nei with gamma distributed rates ($\alpha = 0.437$, pinvar = 0.464). The ML tree is shown in Fig. 1 (excluding outgroups – see above). Three MPT (CI = 0.65, RI = 0.72) were found that agreed in most details with the tree produced by ML, but some important differences were observed. The three MPTs differed from each other only in how the small clade of Australian taxa (*Me. albizziae*, *Au. grevilleae* and *N. exocarpi*) were related. They differed from the ML tree in (1) placing the *Planococcus* species as sister to the *Ferrisia* group, a result found only in the 28S D2 partition in single gene analyses, (2) placing *Vryburgia amaryllidis* as sister to the clade of eight taxa including *Antonina pretiosa* and *Am. lichtensioides* and (3) placing the *Planococcus + Ferrisia* group clade as sister to the clade including *V. amaryllidis*. None of these topological differences were supported by bootstrap analysis however. In general, most nodes were well supported in both ML and MP trees, although weak support for a few nodes indicates that the estimate of phylogenetic congruence could be influenced to some extent by uncertainty in the data.
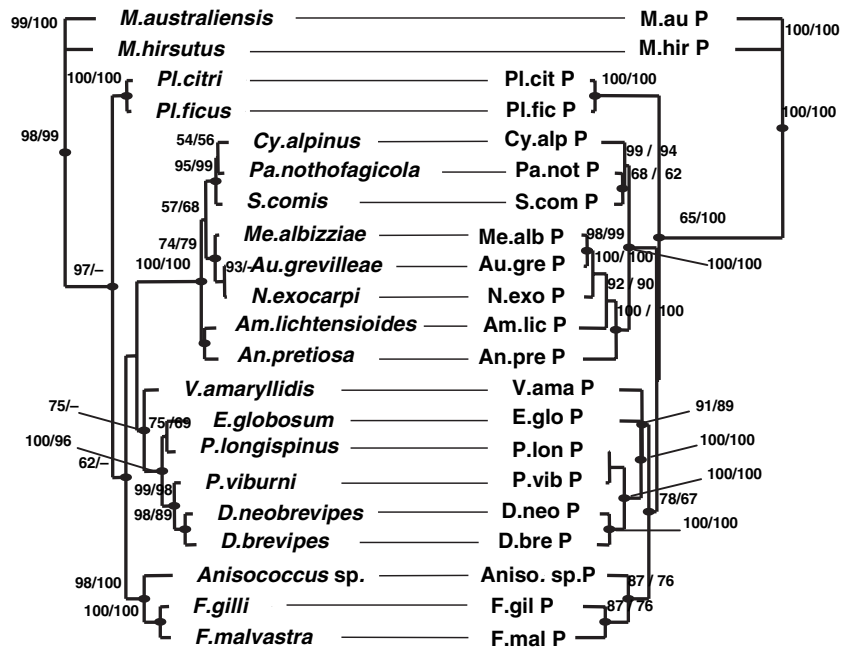
P-endosymbionts
The Tamura-Nei model was used for ML analysis of P-endosymbiont 16S–23S dataset as well ($\alpha = 1.02$, pinvar = 0.347). Deletion of the P-endosymbiont of *V. brevicruris* from the data set had little impact on the topology or node support of ML or MP trees; the ML tree shown in Fig. 1 differs little from the tree in Thao *et al.* (2002). Both ML and the single MP tree (CI = 0.60, RI = 0.65) did, however, resolve the trichotomy of Thao *et al.*'s clades A + B, C + D, and E with the *Planococcus* P-endosymbionts (clade E) sister to the other two clades. Support for this relationship was better for MP (100%) than ML (65%). In addition, the P-endosymbionts of *V. amaryllidis* and *Erium globosum* were placed as sisters by MP but not by ML (Fig. 1).

Thus, both estimates of phylogeny are reasonably robust, justifying the attempt to estimate phylogenetic congruence between partners.

TreeMap results

For the ML trees 37 reconstructions were found, giving 14 co-divergences but varying from zero to three duplications, seven to 12 sorting events and two to five host shifts (Fig. 2). Fourteen codivergences (of a possible 19) were significantly different from the random expectation (mean = 7.36, SD = 1.59; $P < 0.001$). There was a strong correlation between branch lengths in host and symbiont trees ($r = 0.785$, $P < 0.001$). To minimize the number of host shifts an increasing number of sorting events must be invoked. Thus with only two host shifts (occurring from *Pseudococcus viburni* to *P. longispinus*, which are paraphyletic on all mealybug trees, and from *Am. lichtensioides* to the lineage leading to *Me. albizziae*,

Table 1 Tree statistics for maximum likelihood and maximum parsimony analyses.

Gene fragment	ln likelihood	No. of MPTs	L	CI	RI	HI
Mealybugs						
18S	−1763.17	276	170	0.718	0.726	0.410
D2	−1713.28	3	423	0.697	0.783	0.349
D10	−2749.11	1065	344	0.736	0.643	0.265
EF-1α	−1894.16	6	318	0.535	0.585	0.465
Combined	−8390.07	3	1155	0.650	0.723	0.443
P-endosymbionts	−18396.53	1	2657	0.756	0.650	0.244

MPTs, most parsimonious trees.

Fig. 1 Maximum likelihood trees based on the combined data sets for the mealybugs (left) and the 16S–23S rDNA of their P-endosymbionts (right). Bootstrap values are for 200 replications for maximum likelihood (ML) and 500 replications for maximum parsimony (MP) (ML/MP). Outgroups are excluded from the figure because of the long branch subtending the bacterial outgroup taxa. The two *Maconellicoccus* are sister species for both host and symbiont trees. Mealybug names are italicized, P-endosymbionts are in roman font, with specific epithets abbreviated and followed by P. Abbreviations for the genus names of the mealybugs are as follows: *Am, Amonostherium; An, Antonina; Aniso, Anisococcus; Au, Australicoccus; Cy, Cyphonococcus; D, Dysmicoccus; E, Erium; F, Ferrisia; M, Maconellicoccus; Me, Melanococcus; N, Nipaecoccus; P, Pseudococcus; Pa, Paracoccus; Pl, Planococcus; S, Sarococcus; V, Vryburgia.*

N. exocarpi and *Au. grevilleae*) 12 sorting events must be postulated, whereas with five host shifts as few as seven sorting events can explain the data.

Maximum parsimony was not better than ML at recovering a history of perfect phylogenetic congruence – additional incongruence was introduced by the placement of the *Planococcus* species, the *Ferrisia* group, and *V. amaryllidis* noted above, although the MPTs were congruent with the P-endosymbiont tree in the relationships of *An. pretiosa* and *Am. lictensioides*. For the full data set only 13 codivergences are indicated in TreeMap analyses for two of the three mealybug MPTs (Table 2), and a minimum of two to a maximum of seven host shifts are required. The third MP tree reconciles the clade of [*N. exocarpi* + (*Me. albizziae* + *Au. grevilleae*)] with the P-endosymbiont tree producing 14 codivergences. These results indicate that uncertainty in the estimation of relationships among these three taxa affected estimation of phylogenetic congruence. Resolving their relationships did not alter the inference of a host shift onto the line leading to them, however. Interestingly, many more optimal reconstructions were found with MP than ML trees, although this difference was less with MP tree 2 (Table 2). The differences between ML and MP topologies indicate that inferences of phylogenetic congruence may be affected by method of phylogenetic analysis

(although the underlying data may cause the methods to differ).

SOWH test results

The null distribution of δ included values spanning the range 1.632–3.221. The observed δ was 36.552, well outside the null distribution. We can conclude from this result that the observed incongruence between mealybug and P-endosymbiont trees was not caused by random sampling error in the data. We cannot extend this conclusion to other causes of data set incongruence however (see below).

Character congruence and partitioned Bremer support

The ILD tests contrasting mealybug and P-endosymbiont data sets indicated significant incongruence for all comparisons except that with the mealybug 18S gene (all $P < 0.05$, Bonferroni corrected for multiple comparisons).

Given that there is evidence for some incongruence between mealybug and P-endosymbiont trees, PBS could provide insight into the nature of the character conflict. Four MPTs were found from the combined analysis of all
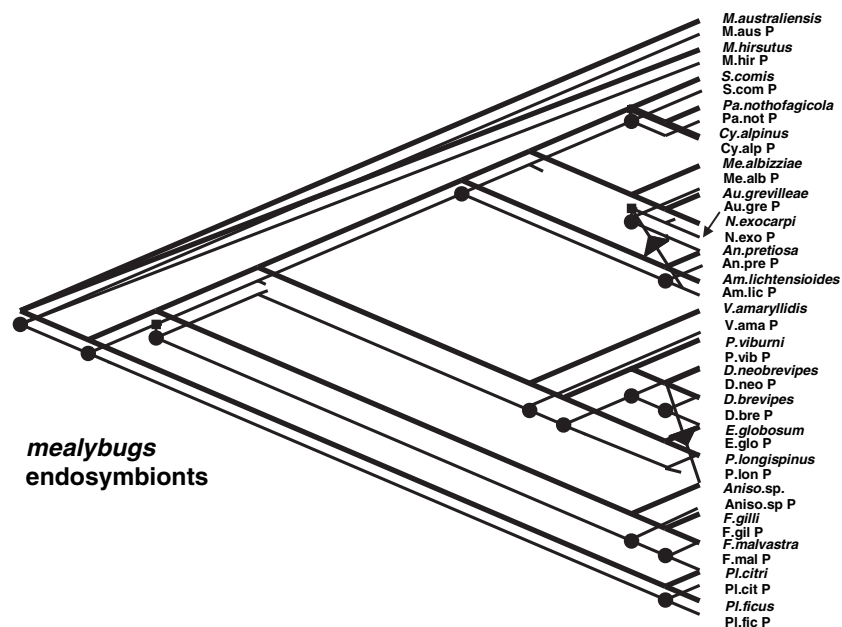
Fig. 2 One of 37 reconstructions of coevolutionary history maximizing the number of codivergences between mealybugs and P-endosymbionts. This reconstruction has the minimum number of inferred host shifts. Trees were based on the combined mealybug data, and the 16S–23S rDNA of the P-endosymbionts. Circles represent codivergence (14), squares duplications (3), short truncated lines sorting events (5), and arrows host shifts (2). The mealybug tree has thick shaded lines and the symbiont tree has thinner bold lines, which are displaced below the host tree so that associated nodes are in the same vertical plane. Accommodation of duplications and sorting events while keeping nodes in the same plane causes terminals to appear mis-matched in some cases. Mealybug names are italicized, P-endosymbionts are in boldface, with specific epithets abbreviated and followed by P.

Table 2 Summary of results from TreeMap analyses for both maximum likelihood (ML) and all three most parsimonious trees (MPT) using combined mealybug data set.

	ML†	MPT1	MPT2	MPT3
Number of reconstructions	37	1154	364	1154
Number of codivergences	14*	13*	14*	13*
Number of duplications	0–3	0–5	0–5	0–5
Number of sorting events	7–12	9–20	8–17	9–20
Number of host shifts	2–5	2–7	2–7	2–7

*Random expectation ML = 7.36 (±1.59); MPT1 = 7.28 (±1.57); MPT2 = 7.21 (±1.55); MPT3 = 7.21 (±1.58), all $P < 0.001$.
†correlations between symbiont and host branch lengths: ML: $r = 0.785$, $P < 0.001$.

6077 bp, one of which was identical to the P-endosymbiont tree, two of which differed by placing the *Planococcus* species sister to the *Ferrisia* group + (*Pseudococcus* + *Dysmicoccus*) clade, all differing in resolution of the New Zealand taxa, *Pa. nothofagicola*, *Sarococcus comis* and *Cyphonococus alpinus*. The strict consensus thus has trichotomies among these taxa (nodes one and 16, Fig. 3). Recovering a tree identical to the P-endosymbiont tree is not surprising as this data set is much larger and has nearly 2.5 times as many parsimony informative characters (weighting the mealybug data by this amount produces a tree closer, but not identical, to the mealybug tree). The PBS values support the overall interpretation of close congruence of the mealybug and P-endosymbiont data sets, although evidence for character conflict appears at nodes 3, 4, 5, 8, 11, 14 and 18 (Fig. 3, Table 3). The conflict is generally weak, but is found in more than one partition in all but one case (node 18), and EF-1α produces substantial conflict at node eight. At four of these nodes (3, 11, 14, 18) clear evidence for character conflict among mealybug data partitions exists as well as between mealybug and P-endosymbiont partitions.

Discussion

The mutualism between mealybugs and their P-endosymbionts is obligate, and strictly vertical transmission is thought to be the rule (Buchner, 1965; Baumann *et al.*, 1997; Moran & Telang, 1998). Many studies have documented phylogenetic congruence at various taxonomic levels, although most surveyed rather broad swaths of the diversity of hemipterans, and have not shown, but rather assumed, perfect congruence (a substantial proportion of nodes were unresolved). The current study strengthens previous results in a striking way: nearly 75% of nodes on the alternative trees were
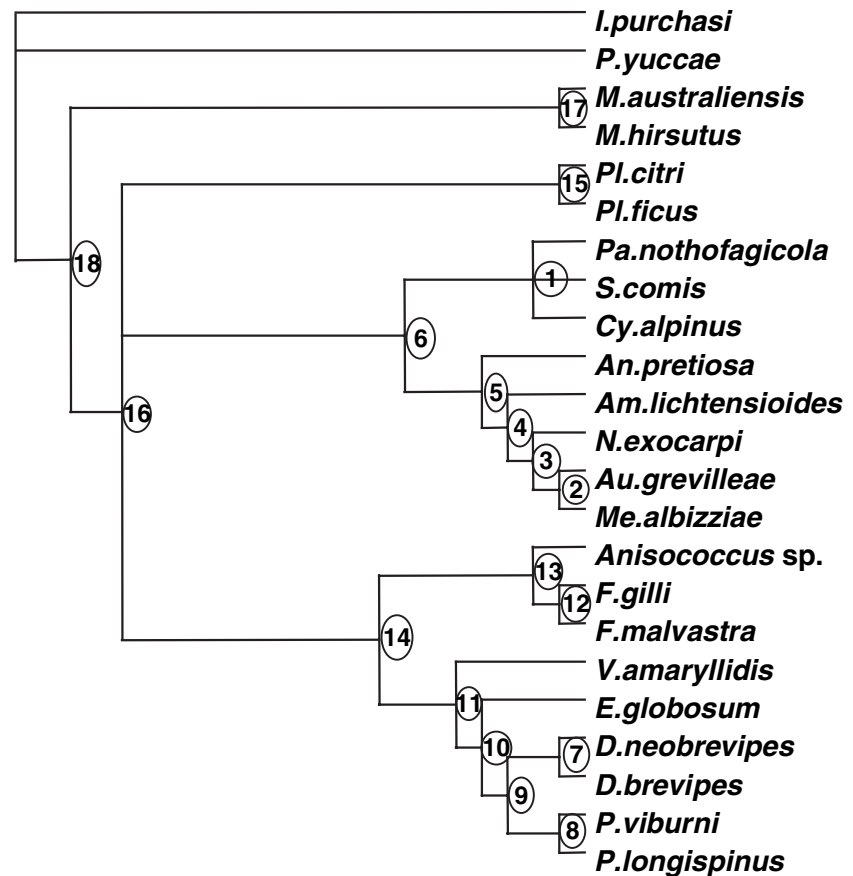
Fig. 3 The strict consensus of four most parsimonious trees from combined analysis of all mealybug and P-endosymbiont data. All characters were given equal weight. Numbers at nodes correspond to node numbers given in Table 3. This tree differs from the P-endosymbiont tree in trichotomies at node 1 and 16.

congruent, and branch lengths of the two trees showed a large and significant correlation. This is strong evidence that codivergence of mealybugs and their P-endosymbionts is the general pattern, as is expected. Of the three possible outcomes other than codivergence (duplications, sorting events and host shifts), the possibility of host shifts is probably the most disturbing to a conception of strict vertical inheritance of endosymbionts. The possibility of a low level of host shifts was unavoidable – for some MP trees it was possible to generate reconstructions lacking host shifts, but these sacrificed a codivergence and were thus sub-optimal (and other incongruent histories were inferred, such as sorting events, which perhaps require more *a posteriori* speculation). The obvious question that needs to be answered is: is this result robust, or is it affected by errors in estimation?

Several potential methodological artefacts may influence any inference of incongruence between host and symbiont trees. First, errors in tree estimation may be inadequately accounted for on a number of levels. Lineage sorting (in the sense of gene trees vs. species trees) may cause discrepancies between the true and estimated trees, particularly involving recently diverged

species, which is where most incongruence is inferred for these data. This process may drastically reduce the probability of recovering congruent histories (Rannala & Michalakis, 2002). Incongruence between different regions of DNA affected by variation in substitution rates may obscure relationships as well. Secondly, the evaluation of confidence in phylogenetic estimates may be misleading. The value of the bootstrap as a confidence measure for the reality of nodes has been criticized (Berry & Gascuel, 1999), and bootstrap values <70% may not reflect a strong level of support in any case (Hillis & Bull, 1993). Thirdly, different methods of tree estimation (MP and ML) may affect the estimation of phylogenetic congruence. All these sources of error may be relevant for some nodes in the mealybug and/or P-endosymbiont trees. Clearly, uncertainty in the estimation of relationships (among *N. exocarpi*, *Me. albizziae* and *Au. grevilleae* for example), data set incongruence (between the EF-1α and 28S D2), and differences in methods of estimation (MP and ML) affected the extent of phylogenetic congruence that was inferred.

On the contrary, results from the SOWH test indicated that the observed incongruence was greater than expected by chance alone, which tends to support the

Table 3 Partitioned Bremer support values for the strict consensus of four most parsimonious trees from four mealybug data partitions and the P-endosymbiont 16S–23S data. The combined analysis tree was based on an equal weights model. The node assignments are shown in Fig. 3.

Node	Data partition					
	18S	28S D2	28S D10	EF-1α	16S–23S	Total
1	1.5	1	2	3	0.5	8
2	0	0	0	0	1	1
3	0.5	−2	−1	3.5	12	13
4	0.5	−3	0	−1	5.5	2
5	0.5	−2	−2	1	7.5	5
6	7.5	16	39	8	10.5	81
7	1.167	4	0.667	3.5	20.667	30
8	0	−1	0	−8	21	12
9	0	0	1	0	22	23
10	−0.5	9	6	0.5	1	16
11	−0.5	5	1	−3.5	3	5
12	2.5	3	5	7.5	40	58
13	4.5	2.5	3.5	3.5	21	35
14	0.5	−4	−2	4	7.5	6
15	1.5	4.5	4.5	1.5	53	65
16	5	5.5	3	4	23.5	41
17	2.5	4	12	0.5	42	61
18	6.5	17	13	−2	284.5	319
Total	33.667	59.5	85.667	26	576.167	781

hypothesis that not all incongruence can be explained by errors in phylogenetic estimation, although the test does not account for the type of systematic error that may accrue through processes such as lineage sorting or rate variation. PBS indicates that character conflict between mealybug and P-endosymbiont genes exists, although in most cases it is quite weak, and conflict exists among mealybug genes as well. It should be noted that because of sampling effects (only a single mealybug sample was collected for each species), the possibility that duplications, sorting events, or host shifts occur cannot be thoroughly tested, and much more intraspecific sampling would be desirable. Results from Funk *et al.* (2001), in the only coevolutionary study of Hemiptera and their endosymbionts in which intraspecific sampling was conducted, suggest that the cohort of symbionts within a host species is likely to be monophyletic, moderating the need for extensive intraspecific sampling. This finding may not hold for all samples, or in all systems.

Errors in phylogenetic estimation may plausibly explain the absence of strict phylogenetic congruence, in a system where it is expected. The implausibility of host shifts on biological grounds tends to lead workers in this field to suspect these inferences. The above arguments do not allow us to go so far as to conclude that congruence must be perfect. Most incongruence can with some confidence be attributed to these causes. Some incongruent nodes were supported as well or better than some congruent nodes in individual as well as combined

data sets however (Fig. 1) – a sister relationship of *P. longispinus* and *P. viburni* was contradicted on all mealybug trees – and host switching is not the only cause of incongruence. Clark *et al.* (2000) were able to use the linkage between two mitochondrial genes to infer that statistically significant phylogenetic incongruence in their data resulted from a violation of the assumptions of the tests resulting in Type I error. They made the claim that because different regions of the nonrecombining mitochondrial genome conflict as much or more with each other as with the *Buchnera* gene *trpB*, it can then be concluded that the statistical tests (KH test and likelihood ratio tests) are rejecting a true hypothesis (the mtDNA must share a common evolutionary history). Despite evidence that aphid mitochondrial genes can be transferred to the nucleus and thus lack a common history (Sunnucks & Hales, 1996), and the possibility of paralogy in nuclear genes such as EF-1α (Hovemann *et al.*, 1988; Danforth & Ji, 1998), their argument seems convincing. We have no such control for the true phylogeny, nor do our data suggest greater incongruence among mealybug genes than between mealybug and P-endosymbiont genes (though incongruence among mealybug genes exists). We therefore cannot reject the hypothesis that host shifts, or evolutionary outcomes other than co-divergence, may occur at a low but discernable frequency. A mechanism for host shifts to occur is not at all clear, and a sorting event implies that re-colonization (a host shift) must have occurred (as all hosts have symbionts currently). Duplications followed by sorting events may not be unexpected.

Although the exchange of endosymbiont bacteria among mealybug species seems unlikely due to the intimate nature of their symbiosis, any host switching is necessarily limited to taxa that are currently sympatric or have been in the geological past. Of the five cases of topology mismatch between the ML trees inferred from mealybug vs. bacterial DNA, three mismatches occur within the tribe Trabutinini (as defined by Downie & Gullan, 2004) for (1) the New Zealand species *Cy. alpinus*, *Pa. nothofagicola* and *S. comis*, (2) the Australian species *Au. grevilleae*, *Me. albizziae* and *N. exocarpi* and (3) for the Nearctic *Am. lichtensioides* and the Palaearctic *An. pretiosa*. The New Zealand species have contiguous or overlapping ranges (Cox, 1987), as do the Australian species (Williams, 1985). Closely related mealybugs are usually geographically related and thus similarity of their endosymbionts could be due either to common ancestry or to recent bacterial interchange. However, for the three mismatches mentioned above, node support in the mealybug gene trees is relatively weak. The other areas of tree mismatch include the Australian species *Erium globosum* and *P. longispinus* (Williams, 1985) and the Nearctic species *P. viburni* (Miller *et al.*, 2002). The endosymbiont data strongly support a monophyletic *Pseudococcus*, whereas traditional study of morphology (Gimpel & Miller, 1996) suggests that members of the *P. maritimus* complex (to

which *P. viburni* belongs) are more closely related to the *Dysmicoccus brevipes* complex than to other *Pseudococcus* species, a result consistent with the mealybug gene tree.

In summary, the data from this study provide perhaps the most convincing example of phylogenetic congruence between interacting partners yet seen, and coupled with the work of Thao *et al.* (2002) reveal a striking history of tandem diversification among interacting partners (mealybugs and endosymbionts). Although this study falls short of demonstrating strict congruence, such a result may be impossible given the limitations of current experimental and statistical procedures, and the small sample of the genome that most studies sequence (see Rokas *et al.*, 2003). Continued research toward innovative methodological approaches to studying the evolutionary history of interacting partners is needed. Nevertheless, at least one example of a host shift was well supported and additional data may reveal that rare instances of violation of strict phylogenetic congruence in the tightest of interactions do indeed occur. Novel associations under the influence of human-mediated transport and mixing of organisms (mealybugs in this case) may lead to novel biologies. Manipulative experiments that attempt to force host shifts in the laboratory could settle the issue. Insect systematists would like to use the symbionts of insect taxa as additional sources of characters for phylogeny estimation. The current study suggests that this approach has merit, but should be used with some caution. Symbionts may be no more prone to homoplasy than other data types, such as morphology or molecules, and may have the advantage of having simpler genomes less prone to mislead because of incongruence among DNA regions seen in the mealybug data set and common in eukaryotic genomes.

Acknowledgments

References

Baker, R.H. & DeSalle, R. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Syst. Biol.* **46**: 654–673.

Baumann, P., Moran, N. & Baumann, L. 1997. The evolution and genetics of aphid endosymbionts. *BioScience* **47**: 12–20.

Berry, V. & Gascuel, O. 1999. On the interpretation of bootstrap trees: approximate threshold of clade selection and induced gain. *Mol. Biol. Evol.* **13**: 999–1011.

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

Clark, M.A., Baumann, L., Muson, M.A., Baumann, P., Campbell, B.C., Duffus, J.E., Osborne, L.S. & Moran, N.A. 1992. The eubacterial endosymbionts of whiteflies constitute a lineage distinct from the endosymbionts of aphids and mealybugs. *Curr. Microbiol.* **25**: 119–123.

Clark, M.A., Moran, N.A., Baumann, P. & Wernegreen, J.J. 2000. Cospeciation between bacterial endosymbionts and a recent radiation of aphids and pitfalls of testing for phylogenetic congruence. *Evolution* **54**: 517–525.

Cook, J.M. & Rasplus, J.-Y. 2003. Mutualists with attitude: coevolving fig wasps and figs. *Trends Ecol. Evol.* **18**: 241–248.

Cox, J.M. 1987. Pseudococcidae (Insecta: Hemiptera). *Fauna of New Zealand* **11**: 1–230.

Danforth, B.N. & Ji, S. 1998. Elongation factor 1-α occurs as two copies in bees: implications for phylogenetic analysis of EF-1-α in insects. *Mol. Biol. Evol.* **15**: 225–235.

Dobzhansky, T. 1937. *Genetics and the Origin of Species.* Columbia University Press, New York.

von Dohlen, C., Kohler, S., Alsop, S.T. & McManus, W.R. 2001. Mealybug β-protobacterial endosymbionts contain γ-protobacterial symbionts. *Nature* **412**: 433–436.

Donaldson, J.S. 1997. Is there a floral parasite mutualism in cycad pollination: the pollination biology *Encephalartos villosus*. *Am. J. Bot.* **84**: 1398–1406.

Downie, D.A. & Gullan, P.J. 2004. Phylogenetic analysis of mealybugs (Hemiptera: Coccoidea: Pseudococcidae) based on DNA sequences from three nuclear genes, and a review of the higher classification. *Syst. Entomol.* **29**: 238–259.

Farris, J.S., Kallerjo, M., Kluge, A.G. & Bult, C. 1994. Testing significance of congruence. *Cladistics* **10**: 315–319.

Funk, D.J., Heibling, L., Wernegreen, J.J. & Moran, N.A. 2000. Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proc. R. Soc. Lond. B* **267**: 2517–2521.

Funk, D.J., Wernegreen, J.J. & Moran, N.A. 2001. Intraspecific variation in symbiont genomes: bottlenecks and the aphid-*Buchnera* association. *Genetics* **157**: 477–489.

Gimpel, W.F. & Miller, D.R. 1996. Systematic analysis of the mealybugs in the *Pseudococcus maritimus* complex (Homoptera: Pseudococcidae). *Contrib. Entomol. Int.* **2**: 1–163.

Goldman, N., Anderson, J.P. & Rodrigo, A.G. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**: 652–670.

Gullan, P.J., Downie, D.A. & Steffan, S.A. 2003. A new pest species of the mealybug genus *Ferrisia* Fullaway (Hemiptera: Pseudococcidae) from the United States. *Ann. Entomol. Soc. Am.* **96**: 723–737.

Hafner, M.S. & Page, R.D.M. 1995. Molecular phylogenies and host–parasite co-speciation: gophers and lice as a model system. *Phil. Trans. R. Soc. Lon. B* **349**: 77–83.

Hillis, D.M. & Bull, J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**: 182–192.

Hillis, D.M. & Huelsenbeck, J.P. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *J. Heredity* **83**: 189–195.

Hovemann, B., Richter, S., Waldorf, U. & Cziepluch, C. 1988. Two genes encode related cytoplasmic elongation factors 1α in *Drosophila melanogaster* with continuous and stage specific expression. *Nucl. Acids Res.* **16**: 3175–3194.

Johnson, K.P., Adams, R.J., Page, R.D. & Clayton, D.H. 2003. When do parasites fail to speciate in response to host specification? *Syst. Biol.* **52**: 37–47.

Karban, R. & Agrawal, A.A. 2002. Herbivore offense. *Annu. Rev. Ecol. Syst.* **33**: 641–664.

Miller, D.R., Miller, G.L. & Watson, G.W. 2002. Invasive species of mealybugs (Hemiptera: Pseudococcidae) and their threat to U.S. agriculture. *Proc. Entomo. Soc. Wash.* **104**: 825–836.

Moran, N.A. & Telang, A. 1998. Bacteriocyte-associated symbionts of insects: a variety of insect groups harbor ancient prokaryotic endosymbionts. *BioScience* **48**: 295–304.

Moran, N.A., Dale, C., Dunbar, H., Smith, W.A. & Ochman, H. 2003. Intracellular symbionts of sharpshooters form a distinct clade with a small genome. *Environ. Microbiol.* **5**: 116–126.

Munson, M.A., Baumann, P., Clark, M.A., Baumann, L., Moran, N.A., Voegtlin, D.J. & Campbell, B.C. 1991. Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. *J. Bacteriology* **173**: 6321–6324.

Page, R.D. 1994. Parallel phylogenies: reconstructing the history of host–parasite assemblages. *Cladistics* **10**: 155–173.

Page, R.D. (ed.) 2002. *Tangled Trees: Phylogeny, Cospeciaton, and Coevolution*. University of Chicago Press, Chicago.

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

Rambaut, A. & Grassly, N.C. 1996. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13**: 235–238.

Rannala, B. & Michalakis, Y. 2002. Population genetics and cospeciation: from process to pattern. In: *Tangled Trees: Phylogeny, Cospeciaton, and Coevolution* (R. D. M. Page, ed.), pp. 120–143. University of Chicago Press, Chicago.

Rokas, A., Williams, B.L., King, N. & Carroll, S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**: 798–804.

Russell, J.A., Latorre, A., Sabater-Munoz, B. & Moran, N.A. 2003. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Molecular Ecology* **12**: 1061–1075.

Schluter, D. 2001. Ecology and the origin of species. *Trends Ecol. Evol.* **16**: 372–380.

Sorenson, M.D. 1999. *TreeRot*, version 2. Boston University, Boston, MA, USA.

Spaulding, A.W. & von Dohlen, C.D. 2001. Psyllid endosymbionts exhibit patterns of co-speciation with hosts and destabilizing substitutions in ribosomal RNA. *Insect Mol. Biol.* **10**: 57–67.

Sunnucks, P. & Hales, D. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion*. *Mol. Biol. Evol.* **13**: 510–524.

Swofford, D.L. 2003. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, version 4.0. Sinauer, Sunderland, MA, USA.

Swofford, D.L., Olsen, G.J., Waddell, P.J. & Hillis, D.M. 1996. Phylogenetic inference. In: *Molecular Systematics* (D. M. Hillis, C. Moritz & B. K. Mable, eds), pp. 407–514. Sinauer Associates, Sunderland, MA, USA.

Thao, M.L. & Baumann, P. 2004. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Appl. Environ. Microbiol.* **70**: 3401–3406.

Thao, M.L., Clark, M.A., Baumann, L., Brennan, E.B., Moran, N.A. & Baumann, P. 2000. Cospeciation of psyllids and their prokaryotic endosymbionts. *Appl. Environ. Microbiol.* **66**: 2898–2905.

Thao, M.L., Gullan, P.J. & Baumann, P. 2002. Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. *App. Env. Micro.* **68**: 3190–3197.

Thompson, J.N. 1994. *The Coevolutionary Process*. University of Chicago Press, Chicago.

Walling, L. 2000. The myriad plant responses to herbivores. *J. Plant Growth Regul.* **19**: 196–216.

Williams, D.J. 1985. *Australian Mealybugs*. British Museum (Natural History), London, UK.

Wu, C.-I. 2001. The genic view of the process of speciation. *J. Evol. Biol.* **14**: 851–865.