Revisiting the first case of insect-bacteria cospeciation: phylogenetic incongruence between aphids and their obligate endosymbiont at subfamily level

Lin Liu[1,2,†], Xiaolei Huang[1,†,*], Yuan Wang[1,2], Gexia Qiao[1,*]

[1] Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China;

[2] University of Chinese Academy of Sciences, 19 Yuquan Road, Shijingshan District, Beijing 100049, China

[†]These authors contributed equally to this work.

[*] Correspondence: Dr. Gexia Qiao, Dr. Xiaolei Huang, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China. E-mail: qiaogx@ioz.ac.cn, huangxl@ioz.ac.cn

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.346v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Abstract

It has been widely accepted that aphids and their primary endosymbiotic bacteria *Buchnera* have strictly parallel diversification relationship. As the first reported case of insect-bacteria cospeciation, this parallel diversification hypothesis has been prevalent, in spite of its basis of limited taxonomic sampling and recent doubts. Here we revisit the evolutionary relationships between aphids and *Buchnera* by using much more taxa and genomic data (16S rDNA, ATP synthase β-subunit gene, and gluconate-6-phosphate dehydrogenase gene) to reconstruct the *Buchnera* phylogeny and test its congruence with the host phylogeny. Comparisons of the *Buchnera* phylogeny with morphology- and molecular-based aphid phylogenies indicate phylogenetic incongruence between aphids and *Buchnera* at subfamily level. Current empirical and theoretical evidence indicate two potential mechanisms underlying this incongruence: one is variation in evolutionary rates of *Buchnera* genomes among different aphid lineages; the other is horizontal transmission of *Buchnera* during the radiation of extant aphid subfamilies and tribes from their common ancestor.

Keywords: codiversification; endosymbiont; evolutionary rate; horizontal transmission; phylogenetic relationship.

INTRODUCTION

The coevolution of parasites and hosts has been attracting attention from biologists for decades. The coevolution process is thought to include five different types of events: duplication, failure to speciation, sorting, host switching and cospeciation (Johnson and Clayton, 2004). Of these events, cospeciation occurs between two specialized interactive organisms that have been closely associated during a long period, and consequently, leads to simultaneous speciation of the lineages involved (Page 1994). A central question in the study of cospeciation is the extent of correlation between cladogenesis of host and parasite (Brooks and McLennan, 1991). The topologies of the phylogenies should be congruent if cosepeciation has occurred, which means host and parasite have been parallel diversification. Cospeciation between invertebrate hosts and intracellular bacteria inhabiting specialized host cells have been documented in a number of studies (Baumann *et al*., 2005; Downie and Gullan, 2005). The best known case is probably the aphid-*Buchnera* symbiosis (Moran *et al*., 2008).

Aphids (Insecta: Aphididae), a phloem-sap feeding insect group, are known to have mutualistic associations with a kind of obligate intracellular bacteria, *Buchnera aphidicola* (Proteobacteria: Enterobacteriaceae) (Buchner, 1965; Munson *et al,* 1991a; Baumann *et al*., 2005). The phloem-sap is rich in carbohydrates but deficient in essential amino acids and vitamins (Douglas, 1992). Therefore, the normal development and reproduction of aphids depend very much on *Buchnera*, which supply these essential nutrients to their hosts (Baumann *et al*., 1995; Nakabachi and Ishikawa, 1997). Recent genomic evidence suggests there is extensive metabolite exchange between aphid and *Buchnera*, including sharing of biosynthetic pathways of amino acids and purine (Wilson *et al*., 2010). These bacteria are located in specialized bacteriocytes of organ-like structures in the aphid hemocoel and are maternally transmitted through infection of embryos or eggs (Baumann *et al*., 1995;

Wilkinson *et al*., 2003). A variety of evidence suggest that the association is indispensable to both aphids and *Buchnera*. After treatment with antibiotics, aposymbiotic aphids grow slowly and are unable to reproduce (Sasaki *et al*., 1991). *Buchnera* has never been cultured in artificial media and cannot survive outside of aphids (Douglas, 1992; Baumann *et al*., 2005).

The earliest studies of phylogenetic relationships between aphids and *Buchnera* showed that a 16S rDNA-based phylogeny of *Buchnera* representatives from four aphid lineages (Aphidinae, Chaitophorinae, Eriosomatinae, Mindarinae; families sensu Heie, 1987; Heie and Wegierek, 2009) was completely concordant with the morphology-based phylogeny of their aphid hosts (Munson *et al*., 1991b; Moran *et al*., 1993). Thus, it was hypothesized that aphids and *Buchnera* have codiversified in parallel since an original infection in the ancestor of modern aphids (Munson *et al*., 1991b; Moran *et al*., 1993). This was referred to as the first case for which codiversification was shown between insects and endosymbiotic bacteria (Moran *et al*., 2008). It was further indicated that this 16S rDNA phylogeny of *Buchnera* was also concordant with the 18S rDNA phylogeny of the corresponding aphid hosts (Baumann *et al*., 1997). Based on ages of aphid fossils and biogeographical events as well as estimated substitution rates of 16S rDNA, the minimum age of the endosymbiotic association between aphids and *Buchnera* was estimated at 160-280 mya (Moran *et al*. 1993).

This codiversification hypothesis was proposed based on 16S rDNA-based *Buchnera* phylogeny by using maximum parsimony (MP) analysis (Munson *et al*., 1991b; Moran *et al*., 1993). However, a 16S rDNA phylogeny of *Buchnera* using maximum likelihood (ML) method with similar taxon sampling yielded a different tree topology, in which the Aphidinae instead of the Eriosomatinae samples formed a basal clade (van Ham *et al*., 1997). This raises doubts about the validity of the *Buchnera* phylogenies on which the cospeciation hypothesis was based, especially given that much superior model-based methods (ML, Bayesian) are currently available in phylogeny reconstruction. On the basis of molecular data (21 sequences

of 16S rDNA, 12 sequences of the ATP synthase β-subunit) for both aphids from six lineages and their *Buchnera* as well as different tree-building methods (Neighbor-joining, MP and ML), Martìnez-Torres *et al.* (2001) showed that some traditionally accepted phylogenetic relationships for both aphids and *Buchnera* have been challenged.

Given that there are nineteen major subfamilies of extant aphids, as well as six subfamilies with fewer than ten species (Remaudiere and Remaudiere, 1997; Favret, 2013), the breadth of taxonomic sampling in previous studies was limited. This also raises doubts about the validity of the codiversification hypothesis of aphid and *Buchnera* in a broader taxonomic coverage (Huang *et al*., 2011). In spite of potential defects, this hypothesis has been prevalent in literature (e.g. Baumann, 2005; Moran *et al*., 2008; Nováková *et al*., 2013). In the past decade, there have been no further empirical studies attempting to test this hypothesis. However, the newest advances in molecular phylogenetics and updated classification of aphids (Heie and Wegierek, 2009; Ortiz-Rivas and Martìnez-Torres, 2010) provide aphid phylogenies which can be used for comparison purpose.

We think that it is time to revisit the phylogenetic relationships between aphids and *Buchnera* as well as the codiversification hypothesis with an improved experimental design. The present study reconstructs the molecular phylogeny of *Buchnera* from major aphid lineages based on expanding taxonomic sampling and increased genomic data, with the specific aims to examine the phylogenetic congruence between aphids and *Buchnera* at higher taxonomic level.

MATERIALS AND METHODS

Taxon sampling, data collection and sequence analysis

To reconstruct a molecular phylogeny of *Buchnera*, 48 aphid species representing 39 genera and 15 major subfamilies were sampled. The number of taxa sampled herein was tripled

93 compared with previous studies which proposed the cospeciation hypothesis. For simplicity
94 and clarity, *Buchnera* was named as their representative aphid hosts. Two bacterial taxa,
95 *Escherichia coli* and *Klebsiella pneumoniae,* which are closely related to *Buchnera*
96 *aphidicola* (Munson *et al,* 1991b), were selected as outgroups. The aphid samples used in this
97 study are listed in Table 1.

98 The specimens for slide-mounting were stored in 75% ethanol, the samples for DNA
99 extraction were stored in 95 or 100% ethanol. All samples and voucher specimens were
100 deposited in the National Zoological Museum of China at the Institute of Zoology, Chinese
101 Academy of Sciences, and kept at -20 °C for long-term storage. Total DNA including the
102 *Buchnera* genome was extracted from single aphid individuals using a standard
103 phenol-chloroform method as well as the DNeasy Blood and Tissue Kit (Qiagen, Germany).
104 Three *Buchnera* genes, the 16S rDNA, the ATP synthase β-subunit (*atpD*) gene, and the
105 gluconate-6-phosphate dehydrogenase (*gnd*) gene, were used in this study. Their sequences
106 were amplified by using published primers, which are listed in Table 2. The PCR conditions
107 for 16S rDNA were as follows: 95℃ for 5 min; 35 cycles consisting of denaturation at 95℃
108 for 1 min, annealing at 65℃ for 0.5 min, and extension at 72℃ for 2 min; and a final
109 extension period at 72℃ for 7 min. PCRs for the ATP synthase β-subunit gene were
110 performed under the following conditions: 94℃ for 5 min; 35 cycles with 94℃ for 1 min, 54
111 ℃ for 0.5 min, and 72℃ for 1 min; followed by 72℃ for 10 min. PCR conditions for *gnd*
112 were as follows: 95℃ for 5 min; 30 cycles with 94℃ for 1 min, 55℃ for 0.5 min, and 72℃
113 for 1 min; and a final 10 min at 72℃. PCR products were visualized by gel electrophoresis in
114 a 1% agarose gel, and products of expected size were cloned into the pMD19-T vector
115 (TakaRa, China). Selected clones were sequenced on an ABI 3730 automated sequencer (ABI,
116 USA). Both strands of the plasmids were sequenced using universal M13 forward and reverse

primers. At least two clones for each sample were sequenced.

Raw sequences were imported into the SeqManII module of Lasergene v.5.0 (DNASTAR Inc., WI, USA). Default settings were used to remove ambiguous bases and vector sequences. The *Buchnera* sequences were verified by using the NCBI BLAST. Sequences for the outgroups *E. coli* and *K. pneumoniae* were cited from GenBank: 16S rDNA (accession no. J01859, X87276), *atpD* (accession no. V00267, NZ_ACZD01000039), and *gnd* (accession no. U00096, CP000964). All sequences for each gene fragment were aligned using MEGA 5.05 (Tamura *et al*., 2011) and then verified manually. Properties of the aligned sequences were also investigated using MEGA. DAMBE 5.2.9 (Xia and Xie, 2001) was used to measure substitution saturation in the aligned sequences by calculating the index of substitution saturation, as described by Xia *et al*. (2003).

Phylogenetic analysis of Buchnera sequences

The partition homogeneity test (PHT), as implemented in PAUP* 4.0 (Swofford, 2002), was used to test congruence among *Buchnera* data sets. The test was implemented with 100 replicates and 10 random-addition sequences per replicate. All data sets with different gene and codon position partitions were tested. PHT test results of the combined data sets suggested some incongruence (P value=0.01, see Table S2). However, the PHT test may suffer from type I errors (e.g. Hipp *et al*., 2004). Besides, ML and Bayesian methods allow data partitions and estimating respective model parameters during the analyses for combined data. Therefore, we compared the topologies of individual genes to each other and to the combined data sets.

Given that we obtained fewer sequences of *atpD* gene than the other two (see Results section), two combined data sets were created. One (B1) concatenated the sequences of 16S rDNA and *gnd* in order to include more taxa, the other (B2) concatenated the sequences of all

three genes. For both of these combined datasets, third codon positions were removed from the *gnd* and *atpD* sequences because of saturation of nucleotide substitutions (see Table S1). Phylogenetic analyses were performed on the individual data sets of the three genes and the two combined data sets, respectively, using MP, ML and Bayesian Markov Chain Monte Carlo (MCMC) methods. Nucleotides were treated as unordered characters with four alternative states, and gaps were considered as missing data in all analyses.

jModelTest v.0.1.1 (Posada, 2008) was used in combination with PAUP*4.0 to select appropriate nucleotide substitution models for all data sets in subsequent ML analyses. The models were chosen according to Akaike information criterion (AIC). The best-fit model for *gnd*, *gnd*1st+2nd and *atpD* genes was TVM + I + G, for *atpD*1st+2nd was TVMef + I + G, and GTR + I + G was the best-fit model for 16S rDNA and the combined data sets. ML analyses were performed in RAxML (Stamatakis *et al*., 2006), under random starting trees with the GTRGAMMA substitution model for individual gene and for each partition of combined datasets. All the parameters were estimated during the ML analyses. Topological robustness was investigated using 1000 non-parametric bootstrap replicates. To compare with the results of RAxML, we also ran PhyML 3.0 (Guindon *et al*., 2010) and Garli 0.96 (Zwickl, 2006). The ML analyses in PhyML applied the optimal substitution models obtained from jModelTest and estimated all model parameters. Branch support was assessed by bootstrap analysis with 1000 replicates. Partitioned ML analyses in Garli were executed on each partitioned multi-gene dataset, with each partition using its optimal model, unlinking and employing their own rates. All other parameters in the control file were kept as default. Clade support values were estimated with 100 bootstrap replications.

Bayesian analyses were performed on all data sets using MrBayes 3.0 (Ronquist and Huelsenbeck, 2003). The best fitting models for each analysis were generated from jModelTest results. The combined data sets were partitioned into different gene partitions, and

each partition was assigned an appropriate nucleotide substitution model according to the jModelTest results. The values of model parameters were treated as unknown variables with uniform prior probabilities, and were estimated during the analyses. For each data set, one cold and three heated MCMC chains were run for one million generations, with trees sampled every 100 generations, and starting from a random tree. MCMC runs were repeated four times to test whether the chains provided valid samples from the posterior probability distribution. The first 2500 trees (25%) for each data set were discarded as burn-in, and the remaining trees were used to construct Bayesian consensus trees. An examination of the log likelihood scores and the average standard deviation of split frequencies (less than 0.01) among parallel chains suggested that the burn-in periods were long enough for chains to become stationary.

Aphid phylogeny

We used newest published higher level aphid phylogenies to examine whether congruence exists between aphid and *Buchnera* phylogenies. The cospeciation hypothesis was originally proposed based on a morphology-based aphid phylogeny including four aphid lineages, which was simplified from Heie (1987) (Munson *et al*., 1991b; Moran *et al*., 1993). Recently, Heie and Wegierek (2009) presented an updated morphology-based phylogeny of aphids based on extensive evidence including aphid fossils. The updated phylogeny was almost identical to Heie (1987) except the position of Phloeomyzinae. Therefore, the morphological phylogeny of aphids modified from Heie and Wegierek (2009) and Heie (1987) was used to compare with the *Buchnera* phylogeny with tripled taxa reconstructed herein. This comparison can provide direct evidence for evaluation of the aphid-*Buchnera* cospeciation, which was first proposed based on morphology-based aphid phylogeny.

Several molecular phylogenetic studies of aphids at higher taxonomic levels showed a lack of resolution in deep nodes (von Dohlen and Moran, 2000; Martìnez-Torres *et al*., 2001;

192 Ortiz-Rivas *et al*., 2004) may be due to limited sampling and a rapid radiation of aphid
193 lineages at about late Cretaceous (von Dohlen and Moran, 2000; Huang *et al*., 2012). The
194 molecular phylogeny proposed by Ortiz-Rivas and Martìnez-Torres (2010) sampled more
195 aphid subfamilies and indicated some relatively stable deep relationships (see in Figure S2).
196 This newest molecular phylogeny was also used to compare with the *Buchnera* phylogeny in
197 order to examine more clearly the phylogenetic relationships between aphids and *Buchnera*.
198
199 **Codiversification analysis**
200 For testing codivergence hypotheses between *Buchnera* and aphid, four representative
201 methods were used: TreeMap 1.0 (Page, 1994), Jane 4 (Conow *et al*., 2010),, SH test
202 (Shimodaira and Hasegawa, 1999) and AU test (Shimodaira, 2002).
203 ***TreeMap and Jane***
204 To compare the aphid and *Buchnera* phylogenies and to create reconstructions of the history
205 of aphid-*Buchnera* association, two reconciliation tools were used: TreeMap 1.0 (Page, 1994)
206 and Jane 4 (Conow *et al*., 2010). The analyses were performed with simplified phylogenetic
207 trees, including morphology-based aphid tree modified from Heie and Wegierek (2009) and
208 Heie (1987), molecular-based aphid tree proposed by Ortiz-Rivas and Martìnez-Torres (2010)
209 and simplified *Buchnera* trees reconstructed herein. As TreeMap and Jane analyses required
210 fully resolved tree and 16 rDNA tree showed comb phylogenetic structure, we only chose
211 trees conducted with *gnd*, *gnd*/16 rDNA and *gnd*/16 rDNA/*atpD* combined dataset in the
212 following analyses. The *Buchnera* trees were simplified on principle that the monophyletic
213 branches including several species from the same tribe or subfamily were merged into one
214 clade. According to the requirement of TreeMap that each parasite should have a
215 corresponding host but not every host has to have a corresponding parasite, we deleted as well
216 as merged some terminals from the *Buchnera* trees to fit the simplified morphology- and

molecular-based aphid tree, respectively. For example, there is a clade of Phloeomyzinae in *Buchnera gnd* tree, but not in simplified molecular-based aphid tree. Therefore, we deleted this clade and remained other merged clade as they were.

TreeMap analyses using both exact search and heuristic search were implemented to find optimal reconstructions. Randomization test with 1000 randomly generated endosymbiont trees using Markovian model was run to test the hypothesis that observed number of cospeciation events differed significantly from random expectation.

TreeMap analysis may underestimate host switching in certain cases (Page, 1994). Thus, we used Jane as a complementary tool to test the phylogenetic congruence between aphids and *Buchnera*. Jane analyses were performed with 500 generations and population sizes of 100. Statistical tests were computed using 100 randomizations with random *Buchnera* trees to check for significance of the resolved cost matrix. An edge-based cost model and a node cost model, which were differ in counting the number of cospeciations, duplications and failure to diverge events, were selected to use with default value.

Likelihood ratio tests

To test the null hypothesis that the likelihoods of aphid and Buchnera tree topology do not differ significantly under the same model, we used these tests. If the null hypothesis is rejected, it is assumed that there was some observed incongruence between aphid and Buchnera trees.

We used the SH tests (Shimodaira and Hasegawa, 1999) as implemented in PAUP and AU tests (Shimodaira, 2002) as implemented in CONSEL (Shimodaira and Hasegawa, 2001) to compare the likelihood score of the alternative trees in two ways, based on the Buchnera dataset. First, we used these tests to assess phylogenetic incongruence between *Buchnera* and aphid phylogenies, including morphology-based aphid tree modified from Heie and Wegierek (2009) and Heie (1987), molecular-based aphid tree proposed by Ortiz-Rivas and

Martìnez-Torres (2010) and simplified *Buchnera* tree reconstructed herein (including trees conducted with *gnd*, *gnd*/16 rDNA and *gnd*/16 rDNA/*atpD* combined dataset), on the nucleotide of corresponding *Buchnera* dataset. Second, we used the tests to conduct pairwise comparisons between scores for alternative Buchnera topologies obtained with gnd and the combined B1and B2 dataset.

Test of evolutionary rate among lineages

The codeml program from PAML (version 4.0) (Yang, 2007) was used to estimate nonsynonymous and synonymous rates of *gnd* gene among *Buchnera* different evolutionary lineages. The ML tree conducted from PAUP was selected in this test. The codon frequencies were derived from the average nucleotide frequencies at the three-codon position (F3 × 4 model), and the model was chosen to compute one ratio of nonsynonymous over synonymous substitution (*d*N/*d*S) per branch but fixed for all sites (model = 1 and NSsites = 0). The one-ratio model (model=0) was also conducted to assume an identical dN/dS ratio value for all main branches. Twice the difference in log-likelihood values ($2\triangle \ln L$) between the two models was calculated following a chi-squared (χ^2) distribution with the degrees of freedom equaling the difference in the number of parameter estimated for the model pairs.

RESULTS

Sequence analyses

Forty-one 16S rDNA sequences for 41 species from 14 subfamilies were obtained, with varied length from 1388 to 1424 bp. Forty-two *gnd* sequences for 42 species from 15 subfamilies were amplified, which varied in length from 814 to 933 bp. Twenty-nine *atpD* sequences for 29 species from 13 subfamilies were successfully obtained, with length from 483 to 516 bp. These sequences were deposited in GenBank. Accession numbers are listed in

267 Table 1.

268 The *atpD* sequences were unambiguously aligned. For the quality of phylogenetic analyses,
269 a highly divergent segment of 65 bp in the alignment of *gnd* sequences as well as an
270 ambiguous segment of approximately 290 bp in the 16S rDNA alignment was excluded.
271 Numbers of variable and parsimony-informative positions for the 16S rDNA genes were
272 lower than for the other two genes. The nucleotide composition of both the *gnd* and *atpD*
273 fragments showed an A-T bias (74.3% and 63.4%, respectively). Tests of substitution
274 saturation for all data sets using DAMBE indicated that no obvious saturation could be
275 observed at the first and second codon positions of bacterial genes, and the third codon
276 position for *gnd* was saturated and for *atpD* was useless sequence (Table S1). So we excluded
277 the third codon positions of *atpD* and *gnd* in the following combined datasets analyses.

278 The combined data set B1 was a 1946 bp alignment including 35 ingroup species and two
279 outgroup species. The combined data set B2 concatenating all three genes included 21
280 ingroups and two outgroups, with 2290 bp in length. The characters of the alignments for final
281 analyses of each gene and the combined data sets are presented in Table 3.

282

283 **Phylogenetic analyses**

284 Except *gnd*, in the analyses of the other two single-gene data sets, the branch-support values
285 were insufficient to resolve basal relationships. In contrast, representative monophyletic
286 subfamily clades were all highly supported. Relatively weak support for basal relationships
287 was also revealed by previous studies (e.g. Martìnez-Torres *et al*., 2001). These may reflect
288 the fact that a single gene contains less useful phylogenetic information than a collection of
289 genes.

290 In the *gnd* phylogeny (Figure 1), *Buchnera* from the representatives of subfamily
291 Pterocommatinae was closely related to *Buchnera* from the representatives of Aphidinae. The

292 *Buchnera* from the representatives of Chaitophorinae, Mindarinae, Greenideinae and
293 Hormaphidinae formed a monophyletic group with strong statistical support (Table S2). Two
294 *Buchnera* samples from the paraphyletic Eriosomatinae were placed in a basal position.
295 Although the bootstrap values were slightly low, three clades were generated (Figure 1), i.e.
296 clade A composed of *Buchnera* from representatives of Greenideinae + Calaphidinae +
297 Saltusaphidinae + Phloeomyzinae + Aiceoninae; clade B including *Buchnera* from
298 Chaitophorinae + Hormaphidinae + Neophyllaphidinae + Lachninae + Thelaxinae, and clade
299 C composed of *Buchnera* from Aphidinae + Anoeciinae + Pterocommatinae + Eriosomatinae
300 + Mindarinae.

301 The monophyly of Eriosomatinae also could not be supported in the 16S rDNA trees
302 (Figure S3). For representative subfamily taxa level, results were less conclusive, and branch
303 support values were weaker (Table S2). However, monophyly of Aiceoninae, Greenideinae,
304 Lachninae, Mindarinae and Pterocommatinae was recovered across different analyses and was
305 well supported. Similar to results in Nováková *et al*. (2013), one species of Macrosiphini,
306 *Cavariella salicicola*, grouped with Pterocommatinae in most of 16S rDNA analyses.
307 Comparing to the *gnd* gene, the 16S rDNA phylogenies were with lower resolution for the
308 deepest nodes. For *atpD*, striking lack of phylogenetic structures for taxa levels higher than
309 species was observed, with comb topology, and some species even can not formed a
310 representative monophyletic group. Therefore, we will keep this gene to increase the effective
311 phylogenetic information but will not separately discuss any phylogenetic inferences for this
312 *Buchnera* gene.

313 Similar to the single-gene topologies, all the results from combined datasets B1 (Figure 2)
314 and B2 (Figure 3) placed Pterocommatinae as a sister group to Aphidinae. The monophyly of
315 most of the subfamilies was commonly well supported, contrary to the separate analyses. The
316 *Buchnera* from Eriosomatinae, was almost always paraphyletic with the tribe Eriosomatini

usually placed in a distant position from Fordini and Pemphigini in all the B1 and B2 combined datasets analyses (also present in the results obtained by Nováková *et al*., 2013).

Most of the rooted and unrooted analyses obtained from the B1 combined dataset, which included 16S rDNA and the first and second condon positions of *gnd* genes, showed a topology almost coincidental with the *gnd* analysis. The three main lineages could be recognized except for the unstable position of unstable position of the subfamily Anoeciinae, Calaphidinae and Fordini. The monophyly of *Buchnera* from aphids of Aphidinae, Pterocommatinae, Chaitophorinae, Mindarinae, Greenideinae, Hormaphidinae, Lachninae and Calaphidinae were strongly supported. However, both the BI and ML rooted analyses supported Aiceoninae placed as a basal position and this position was not rejected by the AU and SH tests.

For the combined dataset B2, although only 22 aphid species included in the analyses, most of the topologies obtained were compatible with the three main lineages described above (Figure 3). The position of the representitive tribe Pemphigini appeared to be unstable under different analyses. Some of the topologies supported a basal position for the Pemphigini as *gnd* trees, while some others supported Aiceoninae instead. Although there were some difference among the topologies from *gnd*, B1 and B2, results from AU test and SH test indicated that trees from one dataset are statistically indifferent explanation of the other dataset.

Phylogenetic congruence test

TreeMap and Jane were used to compare the optimal simplified *Buchnera* trees (*gnd*, *gnd*/16S rDNA and *gnd*/16S rDNA/*atpD*, respectively) with both the molecular- and morphology-based aphid trees. The TreeMap results (see in Table S3) showed that there is no strict codivergence occurring among *Buchnera* and their aphid host. Among the six complete

analyses, a total of 1966 different possible reconstructions were produced: 22 reconstructions from morphological aphid-*gnd*, 22 from molecular aphid-*gnd*, 426 from morphological aphid-*gnd*/ *gnd*/16S rDNA, two from molecular aphid-*gnd*/16S rDNA, 1476 from morphological aphid- *gnd*/16S rDNA/*atpD* and 18 from molecular aphid- *gnd*/16S rDNA/*atpD*. None of these reconstructions of events was found to be significant ($p>0.05$). The tanglegrams for the TreeMap analyses are provided in the Figure S4-S6.

Similarly, the p-value of six Jane analyses were greater than 0.05 (Table S4). The analysis of molecular aphid-*gnd*/16S rDNA/*atpD* showed the lowest cost (cost=18) reconstructions. It comprised eight cospeciations, 12 duplications, six host switches and one loss event (Figure S9). Thus, almost none of the analyses found any significant cospeciation. A summary of the results for each of Jane analyses is available in Table S4.

Because the results of TreeMap and Jane didn't indicate a significantly better fit between aphid and *Buchnera* than expected by chance, we tested the hypothesis that the same topology underlines aphid and *Buchnera* phylogenies. The results of both the SH tests and AU tests indicates that there is significant disagreement between the morphological based aphid tree and the *Buchnera* datasets (*gnd*, *gnd*/16S rDNA and *gnd*/16S rDNA/*atpD*). The molecular-based aphid tree topology was also rejected with the *Buchnera* combined data set B1and B2 by the AU test, which the result was significant ($p<0.01$) (Table S6).

Test of evolutionary rate among lineages

The log likelihood value under the free-ratio model, which assumed different dN/dS ratios for different branch, is ι_1= -22529.70. The one-ratio model, which assumes the same ω parameters for the entire tree, leads toι_0= -22422.26. The obvious significant difference between the two models ($2\Delta\iota$=107.44, $p<0.01$) indicating that the dN/dS ratios are indeed different among lineages. Moreover, among *Buchnera* lineages which represented the host

aphid subfamilies or tribes, the substitution rate (dN/dS) appeared a low level of variability (0.0113-0.3083, PAML M1 model). In the following analyses, we excluded d_S values of Aphidinae, Lachninae,, Greenideinae and Calaphidinae, for the synonymous sites of their *gnd* gene reached saturation.

DISCUSSION

Comparision of aphid and *Buchnera* phylogenies

Parallel evolution of aphids and *Buchnera* was proposed based on phylogenetic comparison of limited taxonomic sampling, i.e. only four aphid lineages (Munson *et al*., 1991b; Moran *et al*., 1993). Although phylogenetic investigations of closely related species and intraspecific lineages indicate fine-scale fidelity of parallelism between aphids and *Buchnera* (Clark *et al*., 2000; Jousselin *et al*., 2009; Liu *et al*., 2012), results of some studies have raised doubts about the validity of the parallel evolution hypothesis (van Ham *et al*., 1997; Martìnez-Torres *et al*., 2001). In the present study, the comparisons of the *Buchnera* phylogeny based on more extensive taxon sampling and increased genomic data with morphology- and molecular-based aphid phylogenies of both the tree-based and data-based analyses demonstrate a complete lack of cospeciation signal between *Buchnera* and aphid hosts.

The inferred *Buchnera* and morphology-based aphid trees differ mainly in two main points. First, the relationships among the *Buchnera* samples from Aphidinae and Lachninae differed from morphological schemes. Previous morphological phylogenies (Heie, 1987; Heie and Wegierek, 2009) proposed that these two subfamilies are sister groups. The main reason was that both Aphidinae and Lachninae have the character of dorsoapical capitate tarsal hairs reduced. Our molecular phylogeny of *Buchnera*, however, revealed that they are distantly related, which agrees with the result in Martìnez-Torres (2001) and Nováková *et al*. (2013). Although the bootstrap value was relatively low, the *Buchnera* from the representatives of

Lachninae grouped with Thelaxinae + Chaitophorinae in the *gnd* phylogeny, which gave us a surprise. The *Buchnera* from Aphidinae, Pterocommatinae grouped together, which is also congruent with the result in Nováková *et al*. (2013). Second, Eriosomatinae was polyphyletic in all analyses, which was similar to the previous reports (Martìnez-Torres *et al*., 2001; Nová ková *et al*., 2013). This result is also different from previous morphological-based proposal (Heie, 1987; Heie and Wegierek, 2009).

The newest molecular phylogeny of aphids proposed in Ortiz-Rivas and Martìnez-Torres (2010) supported some relatively stable basal relationships among aphid subfamilies. Three main lineages named A + D (Aphidini, Macrosiphini, Chaitophorinae, Calaphidinae and Pterocommatinae), E + T (Anoeciinae, Eriosomatini, Pemphigini, Hormaphidini, Cerataphidini, Mindarinae and Thelaxinae) and L (Lachninae) were supported in their phylogeny (Figure S2). However, phylogenetic analyses of our molecular data of *Buchnera* show some distinctly different relationships.

In the aphid molecular tree, Anoeciinae and Hormaphidinae tended to be grouped with Thelaxinae. But in our analyses based on B1 combined data sets, all phylogenetic methods supported the existence of a clade including *Buchnera* from Thelaxinae, Chaitophorinae, Calaphidinae and Hormaphidinae, while Anoeciinae clustered together with Lachninae.

Buchnera from Mindarinae clustered with the representatives from Aphidinae and Pterocommatinae in our phylogenies based on *gnd* and *gnd*/16S rDNA/*atpD* combined data sets. This result indicated disagreement with the previous 16S rDNA-based *Buchnera* phylogeny (Muson *et al*., 1991b; Moran *et al*., 1993; van Ham *et al*., 1997) as well as the molecular-based phylogeny of aphids. In the aphid molecular tree (Figure S2), Mindarinae was placed within the E + T lineage. In a phylogeny proposed by Wojciechowski (1992), based on the short gut and the presence of the triommatidium in the first instar larvae, Mindarinae was placed as a sister taxon to a group including subfamilies Anoeciinae,

Eriosomatinae, Hormaphidinae and Thelaxinae.

The position of Lachninae in the aphid phylogeny has been a bone of contention. Although in conflict with the morphological view that Lachninae should be a sister group of Aphidinae proposed by some taxonomists (Heie, 1987; Heie and Wegierek, 2009), recent molecular studies (von Dohlen and Moran, 2000; Ortiz-Rivas *et al*., 2004; Ortiz-Rivas and Martìnez-Torres, 2010) generally support a basal position of Lachninae. However, all *Buchnera* phylogenies based on different data sets and phylogenetic methods in the present study indicate that *Buchnera* from Lachninae are not at the basal place and have closer relationship.

As shown by the TreeMap and Jane analyses, there was no significant cospeciation between the tree topologies of *Buchnera* and their aphid hosts (morphological- and molecular-based aphid trees, repectively). The number of cospeciations are not high, and the optimal reconstructions, duplications, host switches and sorting events may also indicate that aphids and *Buchnera* have no strictly parallel evolution relationship at higher taxonomic levels, even with some cospeciation events. Moreover, a few specimens changing among these *Buchnera* topologies also didn't affect the results of these analyses. Because the majority of the reconstructions were not found to be significant, the patterns and associations of *Buchnera* and aphids are considered to be no different than if they were chosen at random. The following SH tests and AU tests appeared to confirm the significant incongruence from the aspects of statistics.

Phylogenetic incongruence between aphids and *Buchnera* at higher taxonomic levels

It is a challenge to ascertain mechanisms underlying the higher level phylogenetic incongruence between aphids and *Buchnera*. However, based on accumulated evidence, we think there may be two potential biological reasons. One is variation in evolutionary rates of

Buchnera genomes in different aphid lineages due to distinct evolutionary pressure. dN/dS calculations revealed *Buchnera* genome possibly experiencing varying degrees of purifying selection, and the mutations that have occurred independently on the different aphid lineages. Similarly results were just reported in the analysis of Nováková *et al*. (2013). According to the previous studies, evolutionary rate of endosymbionts depend on population structure and the number of cells transferred from mother to offspring. (Rispe and Moran, 2000). The higher dN/dS ratio of *Buchnera* representative Hormaphidinae and Eriosomatinae lineages (0.3083, 0.1133 respectively) indicate that the decreased effectiveness of purify selection appeared when the aphid host population size is smaller or when the number of *Buchnera* cells transmitted in each progeny is small. New evidence show that the evolutionary rates accelerated quicker in *Buchnera* from Lachniae than *Buchnera* from Aphidinae and Eriosomatinae (Pérez-Brocal *et al*., 2011). Considering different aphid lineages usually have their respective life cycles, which may lead to unique population structure, it is very likely that evolutionary rates of *Buchnera* genomes vary among lineages (Rispe and Moran, 2000; Huang et al., 2011). Theoretically, even a slight difference of evolutionary pressure may lead to topological disagreement across long evolutionary time scale. Moreover, phylogenetic incongruence should be more likely to be observed at higher taxonomic levels while phylogenetic congruence observed at lower levels such as closely related species and intraspecific lineages.

The other potential reason is related to the transmission of *Buchnera* among aphid lineages. It has been generally accepted that the obligate *Buchnera* are vertically and maternally transmitted through infection of embryos or eggs (Baumann *et al*., 1995; Wilkinson *et al*., 2003). Vertical transmission has ever been used as a basic reason for strict cospeciation pattern in aphid-*Buchnera* association (Clark *et al*., 2000). This may be true especially for lower taxonomic level phylogenetic congruence between aphids and *Buchnera*. However,

higher level phylogenetic relationships could be more complicated (Huang *et al*., 2011). Even a very low rate of horizontal transfer of symbionts or their genes among host lineages would lead to phylogenetic incongruence across much longer periods of evolutionary time. Although insect symbionts are predominantly transmitted vertically, phylogenetic studies show that most facultative and some obligate symbionts can also be transmitted horizontally (Darby and Douglas, 2003; Jaenike, 2012). No direct evidence for horizontal transmission of *Buchnera* have been found until now, which may be partly due to a reason as Hotopp (2011) stated that the currently prevailing paradigm is to assume vertical inheritance when other evidence is lacking. However, as host switches had found in TreeMap and Jane analyses, it can be inferred that horizontal transmission probably occurred during very early stages of the aphid-*Buchnera* association. It has been proposed that aphid-*Buchnera* association originated from a single occasional infection in a common ancestor of modern aphids (Muson *et al*., 1991b; Moran *et al*., 1993). Phylogenetic studies and fossil evidence indicate that extant aphid lineages are resulted from a rapid radiation at about late Cretaceous (Heie, 1996; von Dohlen and Moran, 2000; Huang *et al*., 2012). Therefore, there might be some opportunities for horizontal transmission of *Buchnera* during the radiation of aphid lineages from their common ancestor. In fact, very rare horizontal transmission events in early stages of evolutionary history can only be revealed by phylogenetic incongruence between endosymbionts and hosts.

Moreover, the genome size of *Buchnera* varies in different aphid lineages (Moran *et al*., 2008). For example, the *Buchnera* genome of a Lachninae species, *Cinara cedri*, is the smallest one of all known *Buchnera* genomes. This *Buchnera* genome has lost most metabolic functions and its symbiont capacity and might be replaced by the highly abundant coexisting secondary symbiont, which indicating the mutualistic relationship between *Buchnera* and its aphid host seems to end up in a replacement (Pérez-Brocal *et al*., 2006; Moran *et al*., 2008).

Not all the living aphid species contain *Buchnera*, some species of the tribe Cerataphidini lack *Buchnera*, and some species produce dwarf males and/or sterile female soldiers may also lack *Buchnera* (Baumann *et al*., 1995). As some loss events were found in TreeMap and Jane, it can be inferred that some representative *Buchnera* may go extinct in the ancient time and different sorting among hosts, which could affect the phylogentic relationships between *Buchnera* and aphid hosts.

In the present study, a prevalent but potentially deficient hypothesis that aphids and their symbiotic bacteria *Buchnera* have codiversified in parallel has been tested with more extensive taxon sampling and increased genomic data. Comparisons of the bacteria phylogeny with the morphology- and molecular-based aphid phylogenies indicated phylogenetic incongruence at higher taxonomic levels. Current evidence suggest two potential biological reasons underlying the phylogenetic incongruence, i.e. variation in evolutionary rates of *Buchnera* genomes exist among different aphid lineages, and horizontal transmission events may have occurred during the radiation of aphid subfamilies and tribes from their common ancestor. Our results indicate the parallel evolution of aphids and *Buchnera* need to be re-considered at different taxonomic levels. Next step will be to analyze the evolutionary properties of bacterial genome among different aphid lineages to better understand the evolution of insect-bacteria symbiosis.

DATA ARCHIVING

Sequences have been deposited at GenBank (accession numbers refer to Table 1).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.346v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

The authors thank Fang Y., Voegtlin D., Yang J.Y., Ma N.B., Jiang J.G., Bai M., Ren S.S., Zhang D., Wang Y.Q. and Lai Y.M. for specimen collection; Yang F.D. for preparing aphid slides. The work was supported by grants from the National Natural Sciences Foundation of China (30830017, 31272348), National Science Funds for Distinguished Young Scientists (31025024), National Science Fund for Fostering Talents in Basic Research (J0930004), and the Ministry of Science and Technology of the People's Republic of China (MOST GRANT No. 2011FY120200).

REFERENCES

Baumann P (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu Rev Microbiol* **59**: 155–189.

Baumann P, Baumann L, Lai CY, Rouhbaksh D, Moran NA, Clark MA (1995). Genetics, physiology and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. *Annu Rev Microbiol* **49**: 55–94.

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

Brooks DR, McLennan DA (1991). Phylogeny, ecology and behavior: a research program in comparative biology. University of Chicago Press, Chicago.

Buchner P (1965). *Endosymbiosis of animals with plant microorganisms*. Wiley Interscience: New York.

Clark MA, Moran NA, Baumann P (1999). Sequence evolution in bacterial endosymbionts having extreme base composition. *Mol Biol Evol* **16**:1586–1598.

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

Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R (2010). Jane: a new tool for the cophylogeny reconstruction problem. *Algorithm Mol Biol* **5**: 16–25.

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

Douglas AE (1992). Requirement of pea aphids (*Acyrtosiphon pisum*) for their symbiotic bacteria. *Entomol Exp Appl* **65**: 195–198.

547 Downie DA, Gullan PJ (2005) Phylogenetic congruence of mealybugs and their primary endosymbionts. *J*
548 *Evol Biol* 18:315–324.
549 Favret C (2013). *Aphid Species File*. Version 1.0/4.1. Retrieval date: 10 November 2013.
550 http://aphid.speciesfile.org.
551 Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and
552 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst*
553 *Biol* **59**: 307–321.
554 Heie OE (1987). Paleontology and phylogeny. In: Minks AK, Harrewijn P (eds) *Aphids, their Biology,*
555 *Natural Enemies and Control*, Elsevier Press: Amsterdam. pp 367–391.
556 Heie OE (1996). The evolutionary history of aphids and a hypothesis on the coevolution of aphids and
557 plants. *Boll Zool Agrar Bachic* **28**: 149–155.
558 Heie OE, Wegierek P (2009). A classification of the Aphidomorpha (Hemiptera: Sternorrhyncha) under
559 consideration of the fossil taxa. *Redia* **2009**: 69–77.
560 Hipp AL, Hall JC, Sytsma KJ (2004). Congruence versus phylogenetic accuracy: Revisiting the
561 incongruence length difference test. *Syst Biol* **53**: 81–89.
562 Hotopp JCD (2011). Horizontal gene transfer between bacteria and animals. *Trends Genet* **27**: 157–163.
563 Nováková E, Hypša V, Klein J, Foottit RG, von Dohlen CD, Moran NA (2013) Reconstructing the
564 phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*.
565 *Mol. Phylogenet. Evol*. 68: 42–54.
566 Huang XL, Liu L, Qiao GX (2011). Evolutionary relationships between aphids and their primary
567 endosymbionts: hypotheses and mechanisms. *Acta Entomol Sin* **54**: 582–588.
568 Huang X, Xiang-Yu J, Ren S, Zhang R, Zhang Y, Qiao G (2012). Molecular phylogeny and divergence
569 times of Hormaphidinae (Hemiptera: Aphididae) indicate Late Cretaceous tribal diversification. *Zool J*
570 *Linn Soc* **165**: 73–87.
571 Jaenike J (2012). Population genetics of beneficial heritable symbionts. *Trends Ecol Evol* **27**: 226–232.
572 Johnson KP, Clayton DH (2004). Untangling coevolutionary history. *Syst Biol* **53:** 92–94.
573 Jousselin E, Desdevised Y, Coeur d'acier A (2009). Fine-scale cospeciation between *Brachycaudus* and
574 *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in aphids.
575 *Proc R Soc B* **276**: 187–196.

576 Liu L, Huang XL, Zhang RL, Jiang LY, Qiao GX (2012). Phylogenetic congruence between
577 *Mollitrichosiphum* (Aphididae: Greenideinae) and *Buchnera* indicates insect–bacteria parallel evolution.
578 *Syst Entomol* doi: 10.1111/j.1365-3113.2012.00647.x
579 Martìnez-Torres D, Buades C, Latorre A, Moya A (2001). Molecular systematics of aphids and their
580 primary endosymbionts. *Mol Phylogenet Evol* **20**: 437–449.
581 Moran NA (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Pro Natl Acad Sci*
582 *USA* **93**: 2873–2878.
583 Moran NA, McCutcheon JP, Nakabachi A (2008). Genomics and evolution of heritable bacterial symbionts.
584 *Annu Rev Genet* **42**: 165–190.
585 Moran NA, Munson MA, Baumann P, Ishikawa H (1993). A molecular clock in endosymbiotic bacteria is
586 calibrated using the insect hosts. *Proc R Soc B* **253**: 167–171.
587 Munson MA, Baumann P, Kinsey MG (1991a). *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a
588 taxon consisting of the mycetocyte associated, primary endosymbionts of aphids. *Int J Syst Bacteriol* **41**:
589 566–568.
590 Munson MA, Baumann P, Clark MA, Baumann L, Moran NA, Voegtlin DJ *et al*. (1991b). Evidence for the
591 establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. *J Bacteriol* **173**:
592 6321–6324.
593 Nakabachi A, Ishikawa H (1997). Differential display of mRNAs related to amino acid metabolism in the
594 endosymbiotic system of aphids. *Insect Biochem Molec* **27**: 1057–1062.
595 Ortiz-Rivas B, Martìnez-Torres D (2010). Combination of molecular data support the existence of three
596 main lineages in the phylogeny of aphids (Hemiptera: Aphididae) and the basal position of the subfamily
597 Lachninae. *Mol Phylogenet Evol* **55**: 305–317.
598 Ortiz-Rivas B, Moya A, Martìnez-Torres D (2004). Molecular systematics of aphids (Homoptera:
599 Aphididae): new insights from the long wavelength opsin gene. *Mol Phylogenet Evol* **30**: 24–37.
600 Page RD (1994) Parallel phylogenies: reconstructing the history of host-parasite assemblages. *Cladistics* **10**:
601 155–173.
602 Pérez-Brocal V, Gil R, Moya A, Latorre A (2011). New insights on the evolutionary history of aphids and
603 their primary endosymbiont *Buchnera aphidicola*. *Int J Evol Biol* Art: 9 ID: 250154.
604 Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM *et al*. (2006). A small microbial

genome: the end of a long symbiotic relationship? *Science* **314**: 312–313.

Posada D (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**: 1253–1256.

Remaudière G, Remaudière M (1997). *Catalogue des Aphididae du Monde*. Institut National de la Recherche Agronomique: Paris.

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

Rispe C, Moran NA (2000) Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *Am. Nat* **156**: 425-441.

Sasaki T, Hayashi H, and Ishikawa H (1991). Growth and reproduction of symbiotic and aposymbiotic pea aphids, *Acyrtosiphon pisum*, maintained on artificial diets. *J Insect Physiol* **37**: 85–92.

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst. Biol. 51*: 492–508.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol* **16**: 1114-1116.

Shimodaira H, Hasegawa M (1999) Consel: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**: 1246-1247.

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

Swofford DL (2002). *PAUP*. Phylogenetic analysis using parsimony (*and other methods)*, version 4. Sinauer Associates, Sunderland, MA.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

van Ham RJ, Moya A, and Latorre A (1997). Putative evolutionary origin of plasmids carrying the genes involved in leucine biosynthesis in *Buchnera aphidicola* (endosymbiont of aphids). *J Bacteriol* **179**: 4768–4777.

von Dohlen CD, Moran NA (2000). Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biol J Linn Soc* **71**: 689–717.

Wilkinson TL, Fakatsu T, Ishikawa H (2003). Transmission of symbiotic bacteria *Buchnera* to

parthenogenetic embryos in the aphid *Acyrthosiphon pisum* (Hemiptera: Aphidoidea). *Arthropod Struct Dev* **32**: 241–245.

Wilson ACC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G *et al*. (2010). Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol Biol* **19**: 249–258.

Wojciechowski W (1992). *Studies on the Systematic System of Aphids* (*Homoptera*: *Aphidinea*). Uniwersytet Slaski: Katowice.

Xia X, Xie Z (2001). DAMBE: data analysis in molecular biology and evolution. *J Hered* **92**: 371–373.

Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. *Mol Phylogenet Evol* **26**:1-7**.**

Yang ZH (2007) PAML4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586-1591.

Zhang HC, Qiao GX (2008). Molecular phylogeny of Pemphiginae (Hemiptera: Aphididae) inferred from nuclear gene EF-1α sequences. *Bull Entomol Res* **98**: 499–507.

Zwickl D (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. University of Texas at Austin; PhD thesis.

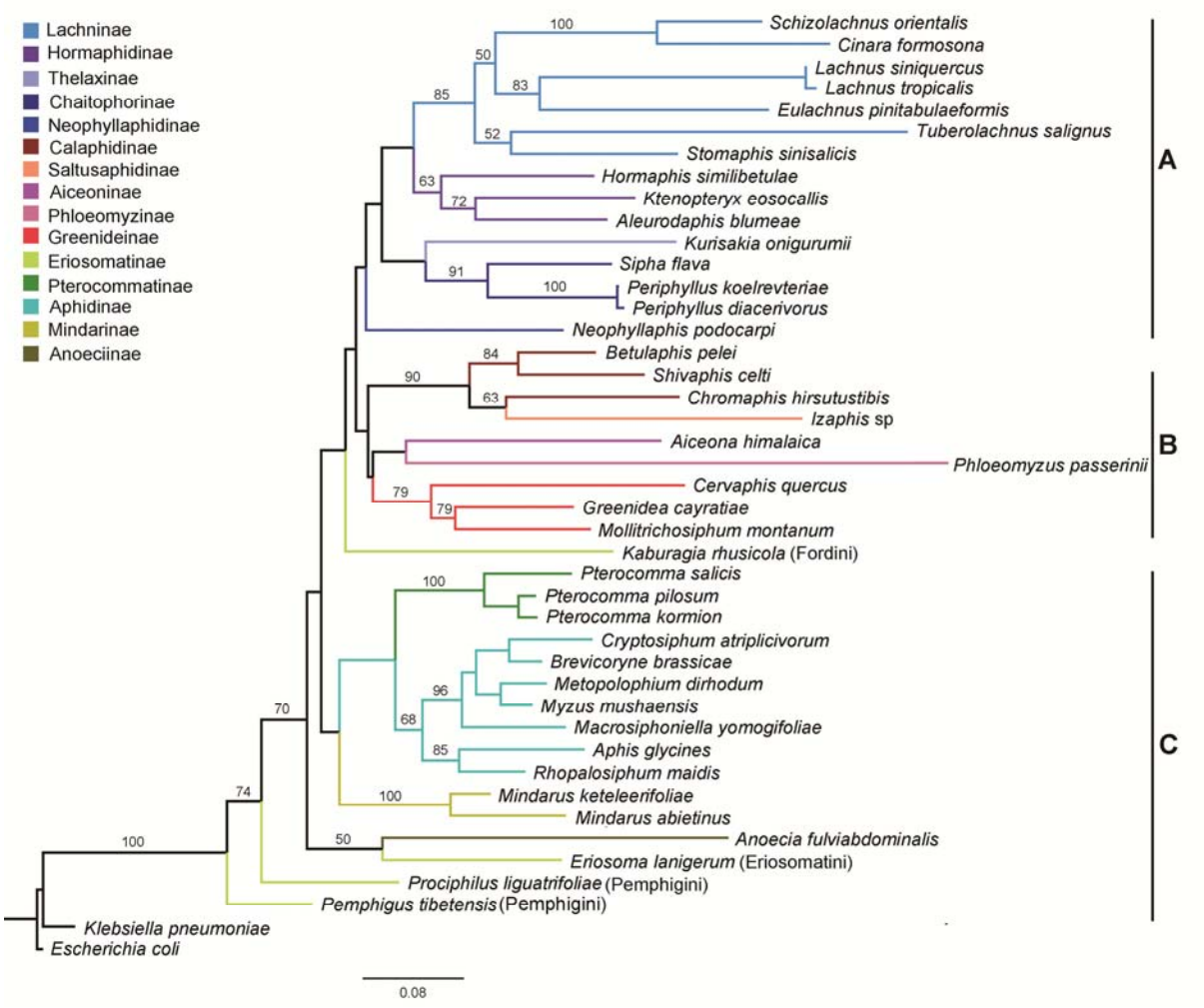

Figure 1 *Buchnera* phylogeny based on maximum likelihood analysis of *gnd* sequences. *Buchnera* samples are indicated as names of their aphid hosts. Values above branches are bootstrap values of maximum likelihood analysis. The bootstrap values higher than 50 are shown.

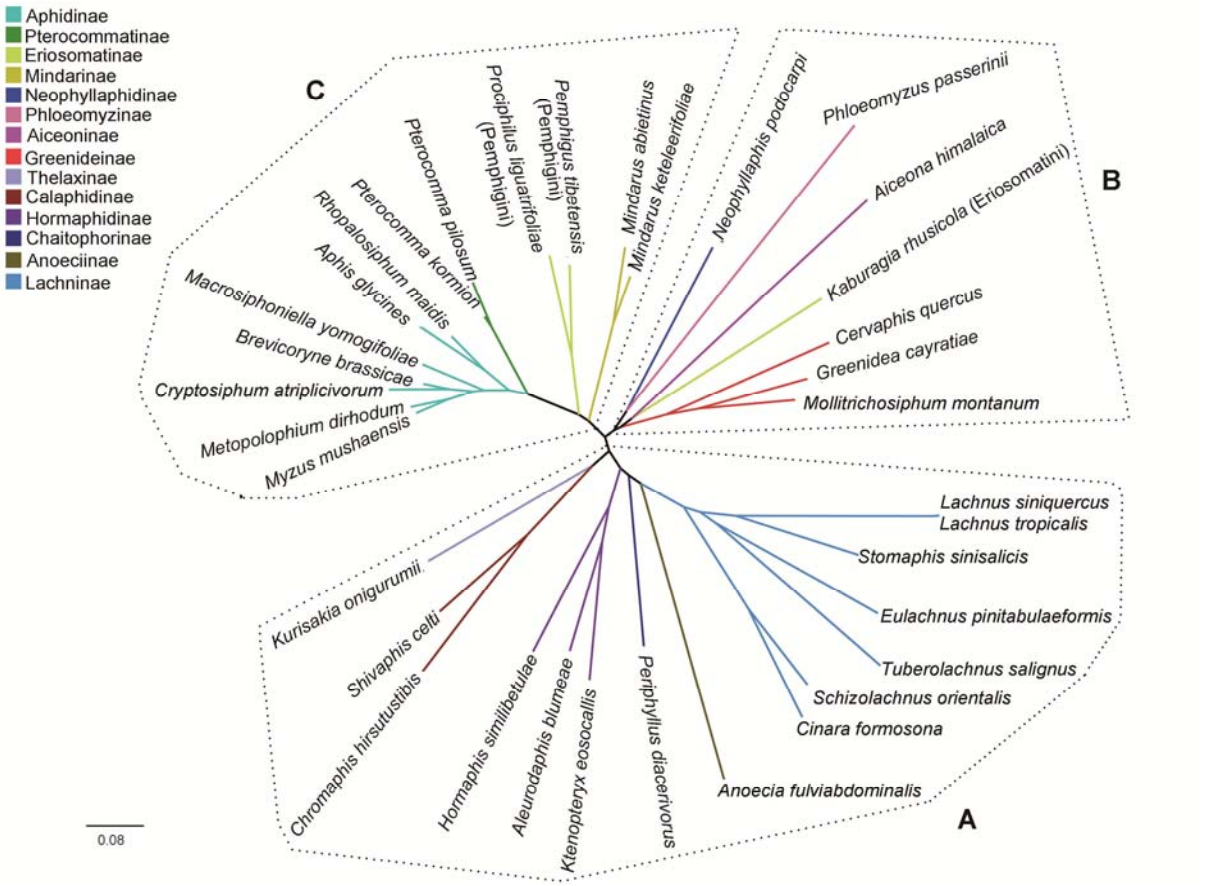

Figure 2 *Buchnera* phylogeny based on maximum likelihood analysis of *gnd*/16S rDNA combined data set. *Buchnera* samples are indicated as names of their aphid hosts. Bootstrap values higher than 50 are shown next to the node. Dashed ellipses enclose the three main lineages A, B and C (see text).

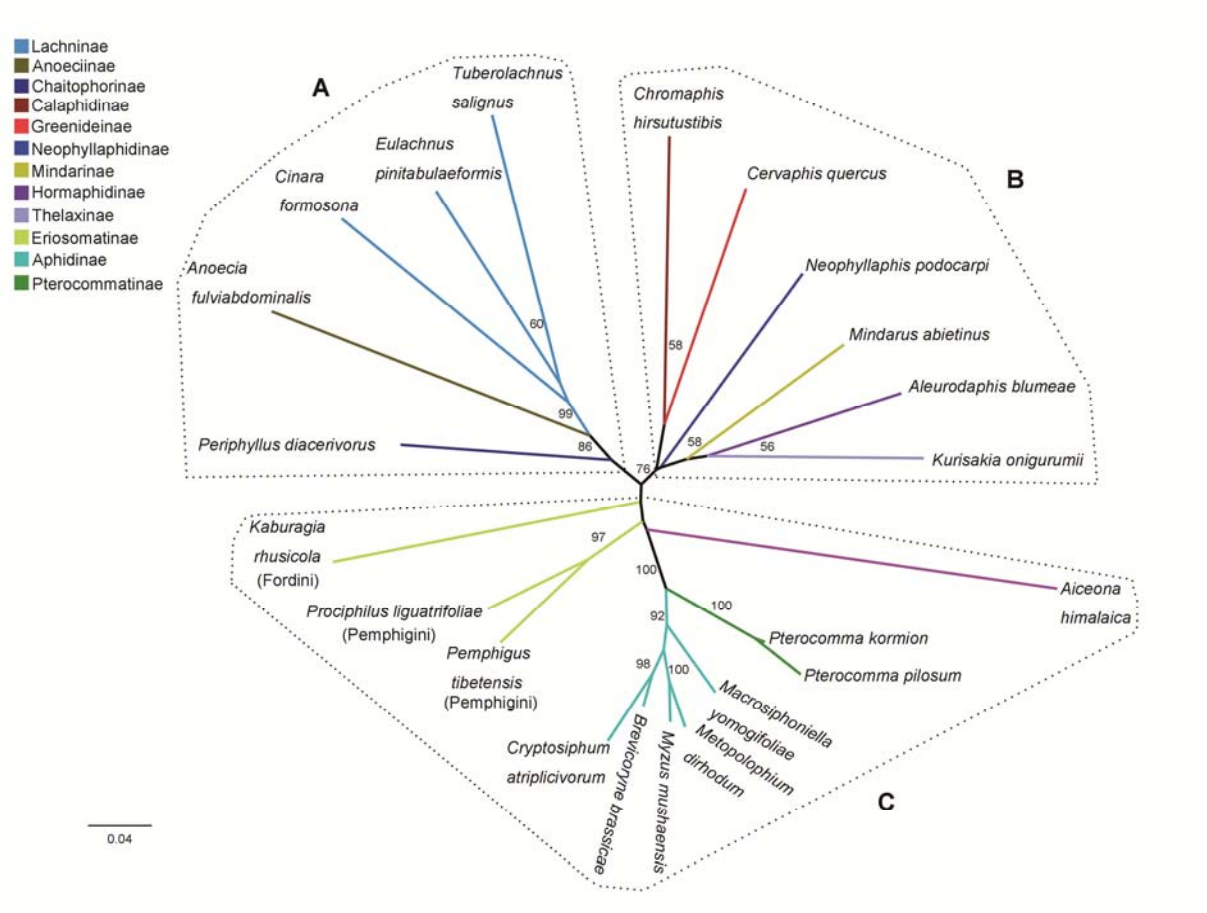

Figure 3 *Buchnera* phylogeny based on maximum likelihood analysis of *gnd*/16S rDNA/*atpD* combined data set. *Buchnera* samples are indicated as names of their aphid hosts. Bootstrap values higher than 50 are shown next to the node. Dashed ellipses enclose the three main lineages A, B and C (see text).

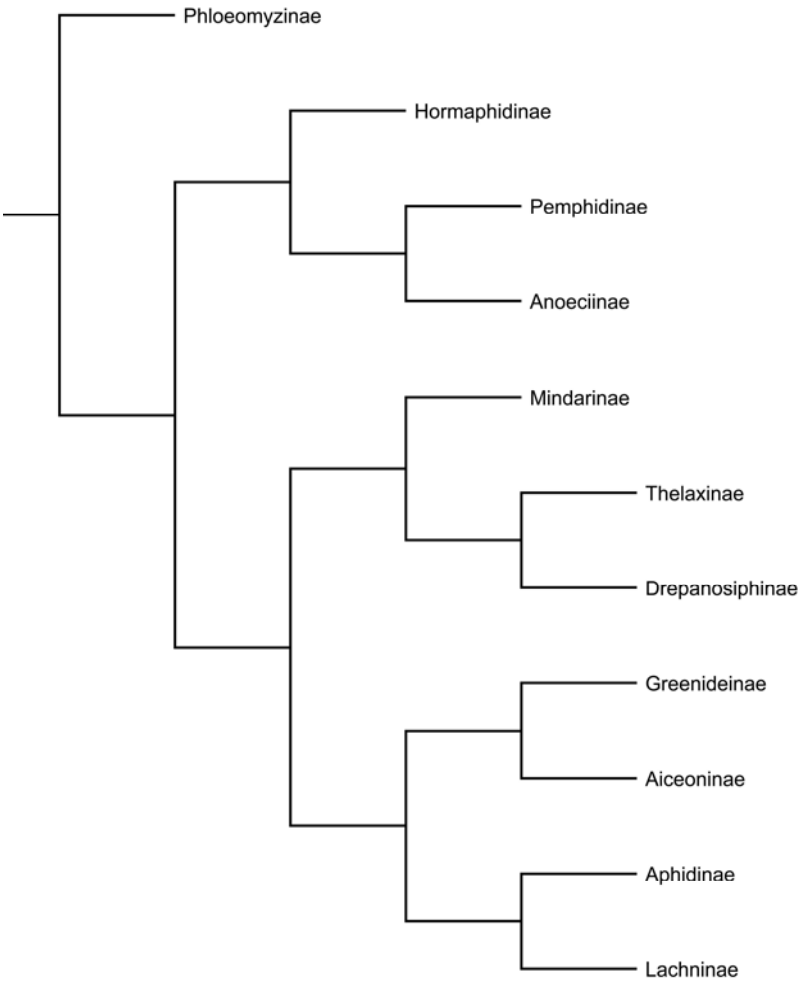

Figure S1 Morphology-based aphid phylogeny modified from Heie (1987) and Heie and Wegierek (2009). We use the subfamily-level classification system of Remaudière and Remaudière (1997), in which Heie's Drepanosiphidae is divided into Chaitophorinae, Calaphidinae, Neophyllaphidinae and Saltusaphidinae; Aphididae is divided into Aphidinae and Pterocommatinae.

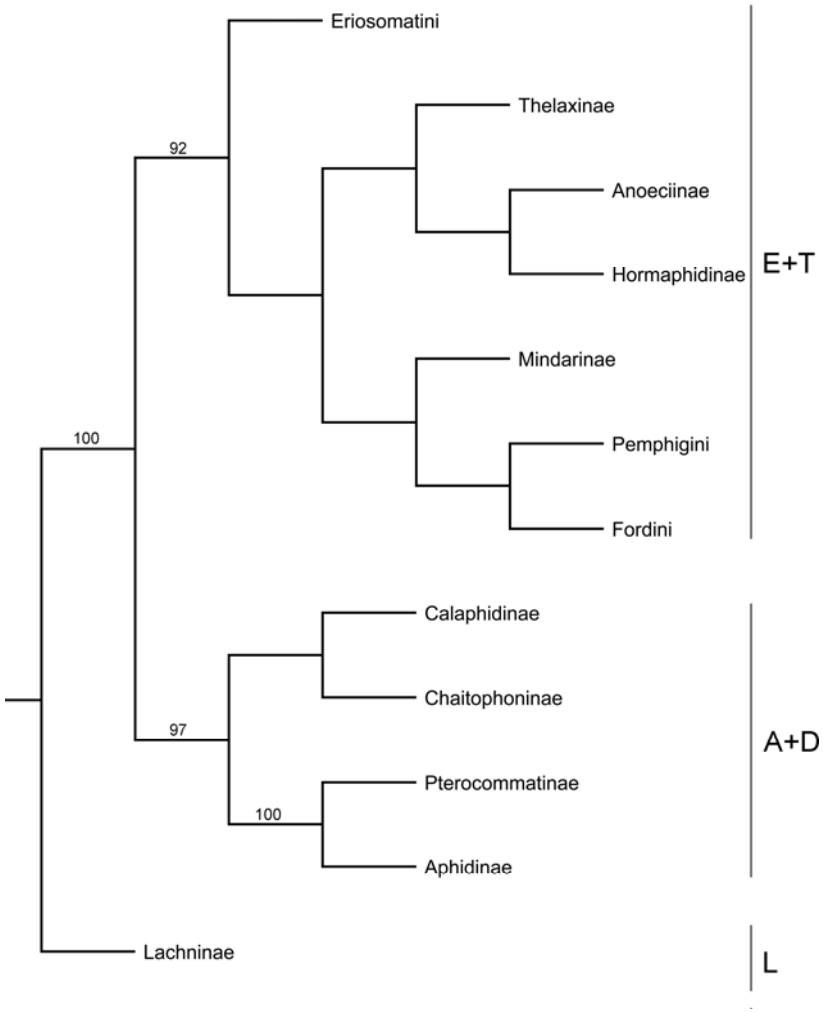

Figure S2 Molecular-based aphid phylogeny modified from Ortiz-Rivas and Martìnez-Torres (2010). Only those subfamilies included in the present report are shown.

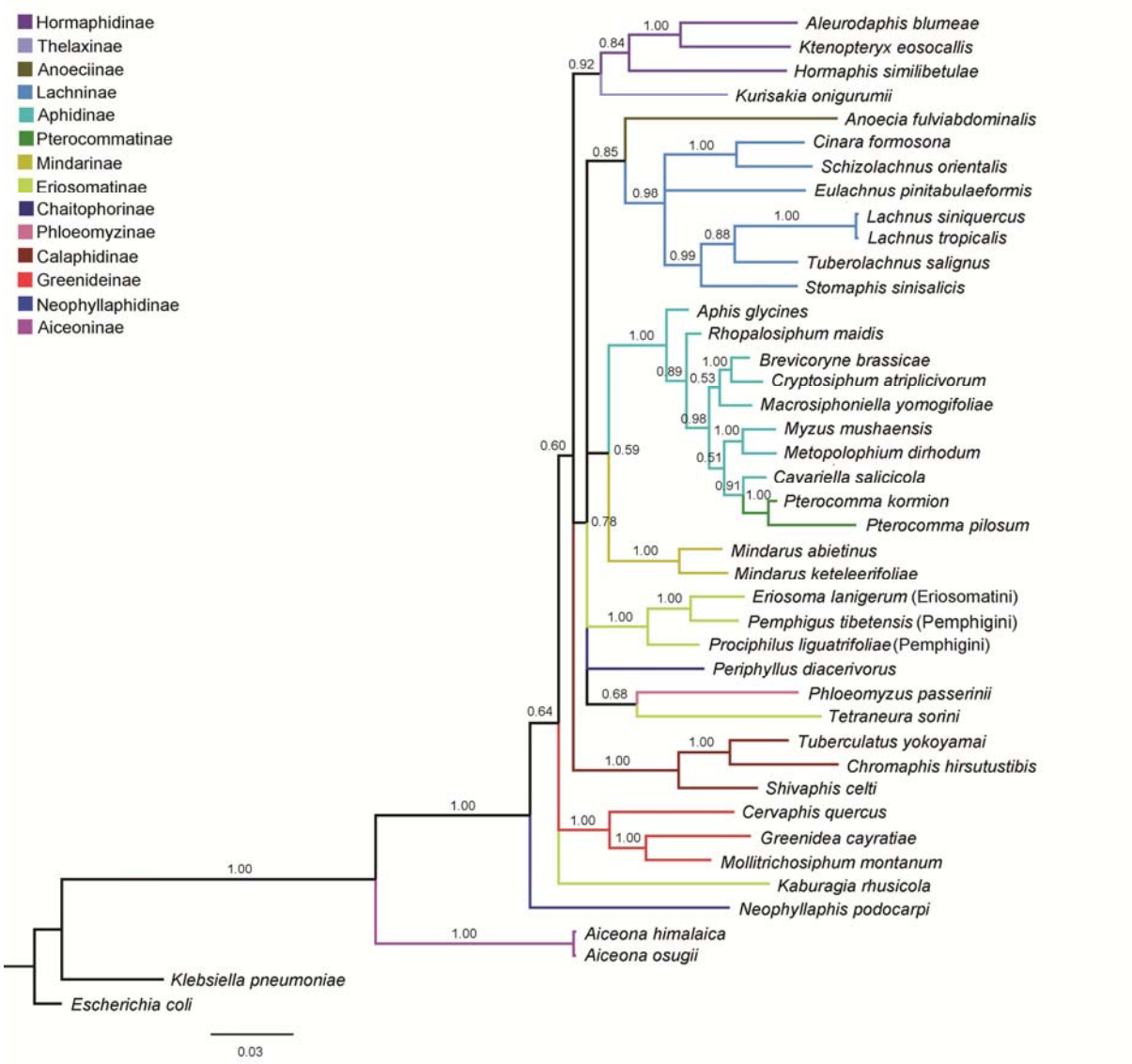

Figure S3 *Buchnera* phylogeny based on Bayesian analysis of 16S rDNA sequences. *Buchnera* samples are indicated as names of their aphid hosts. Values above branches are posterior probablities of Bayesian inference. The bootstrap values higher than 50 are shown.

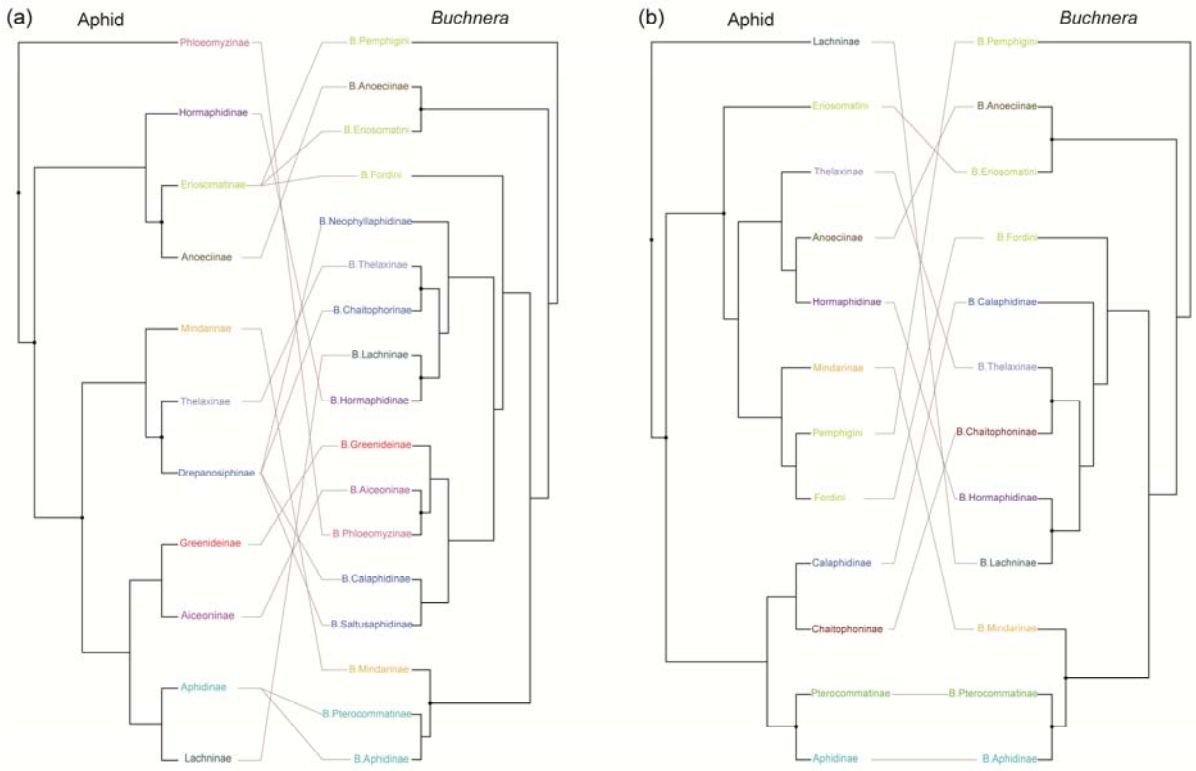

Figure S4 TreeMap comparison of aphid and *Buchnera* phylogenies. (a) The left is the morphological-based aphid phylogeny simplified from Heie (1987) and Heie and Wegierek (2009); the right is simplified *gnd Buchnera* phylogeny with outgroups omitted. (b) The left is the molecular-based aphid phylogeny simplified from Ortiz-Rivas and Martìnez-Torres (2010); the right is simplified *gnd Buchnera* phylogeny. All the analyses are omitted outgroups.

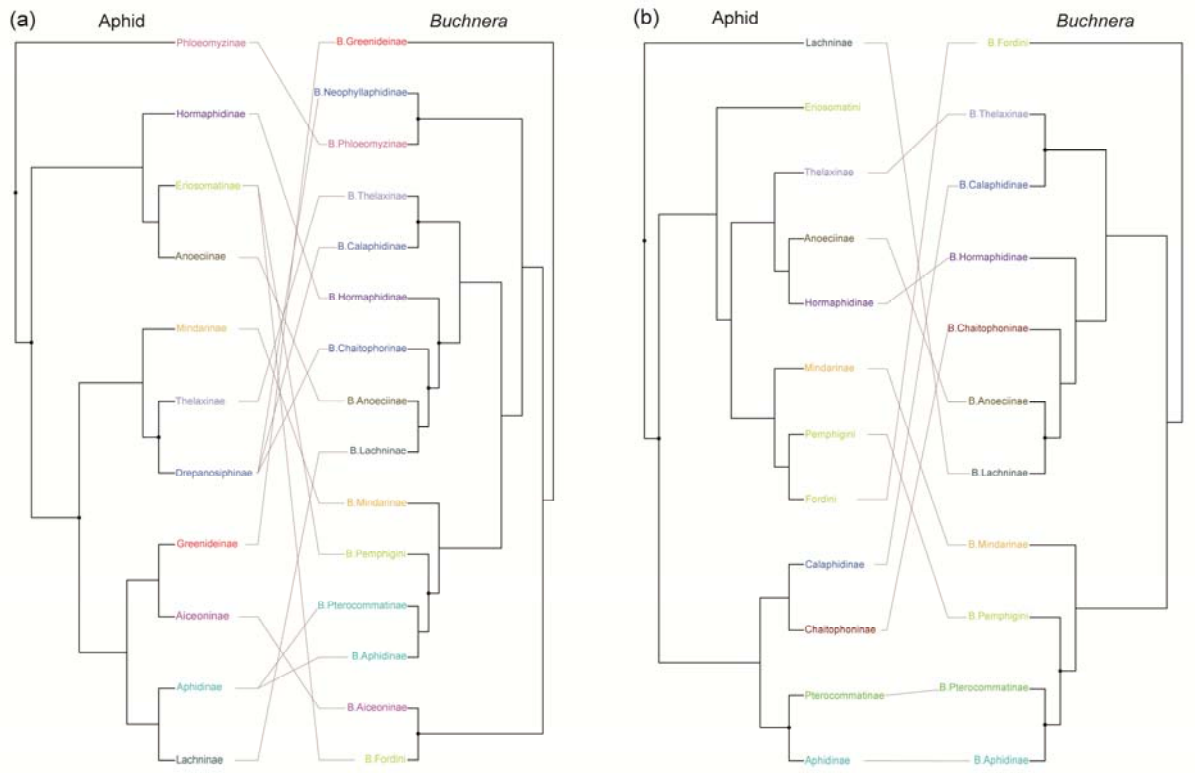

Figure S5 TreeMap comparison of aphid and *Buchnera* phylogenies. (a) The left is the morphological-based aphid phylogeny simplified from Heie (1987) and Heie and Wegierek (2009); the right is simplified *gnd*/16S rDNA *Buchnera* phylogeny with outgroups omitted. (b) The left is the molecular-based aphid phylogeny simplified from Ortiz-Rivas and Martìnez-Torres (2010); the right is simplified *gnd*/16S rDNA *Buchnera* phylogeny. All the analyses are omitted outgroups.

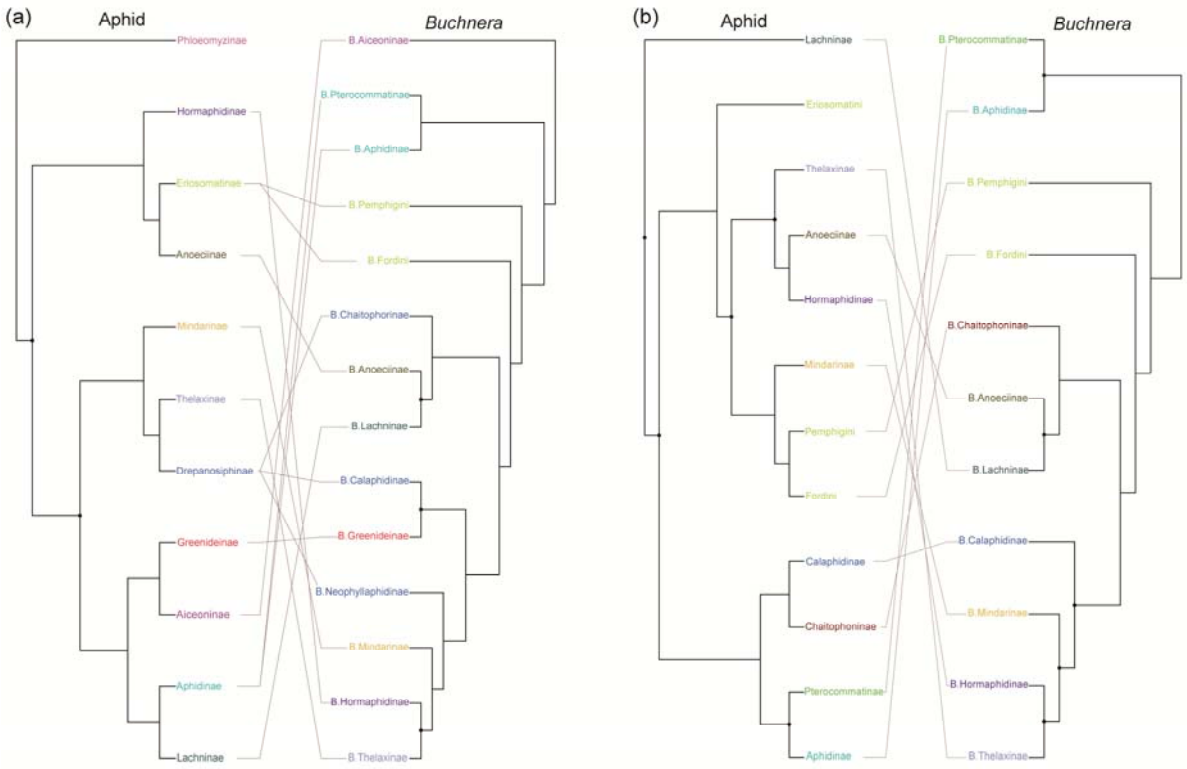

Figure S6 TreeMap comparison of aphid and *Buchnera* phylogenies. (a) The left is the morphological-based aphid phylogeny simplified from Heie (1987) and Heie and Wegierek (2009); the right is simplified *gnd*/16S rDNA/*atpD Buchnera* phylogeny with outgroups omitted. (b) The left is the molecular-based aphid phylogeny simplified from Ortiz-Rivas and Martìnez-Torres (2010); the right is simplified *gnd*/16S rDNA/*atpD Buchnera* phylogeny. All the analyses are omitted outgroups.

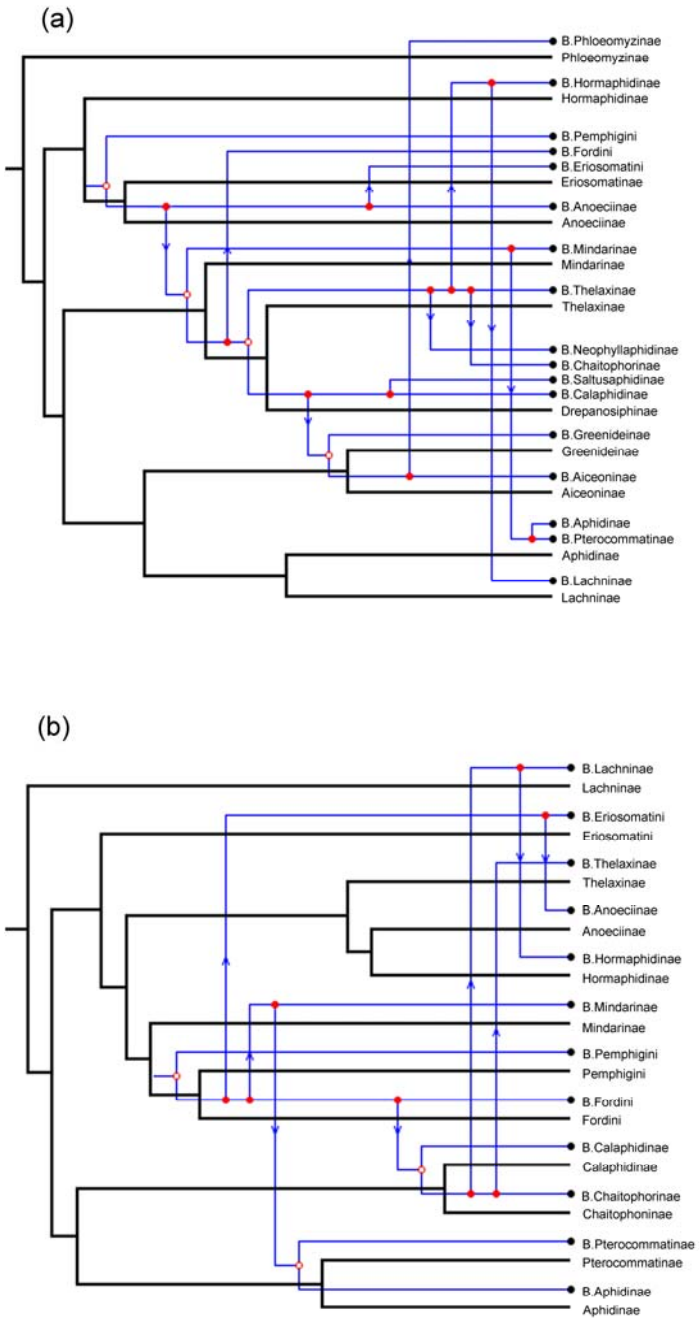

Figure S7 Cophylogeny of aphid and *Buchnera* from Jane. (a) The reconciled trees based on the morphology-based aphid tree and the simplified gnd *Buchnera* tree. (b) The reconciled trees based on the simplified molecular-based aphid tree and the simplified *gnd Buchnera* tree. Blue and black lines indicate the phylogenies of the *Buchnera* and aphids, respectively. Hollow red circles indicate cospeciation events; solid red circles indicate duplications; arrows indicate host switch events; dotted lines indicate loss events.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.346v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

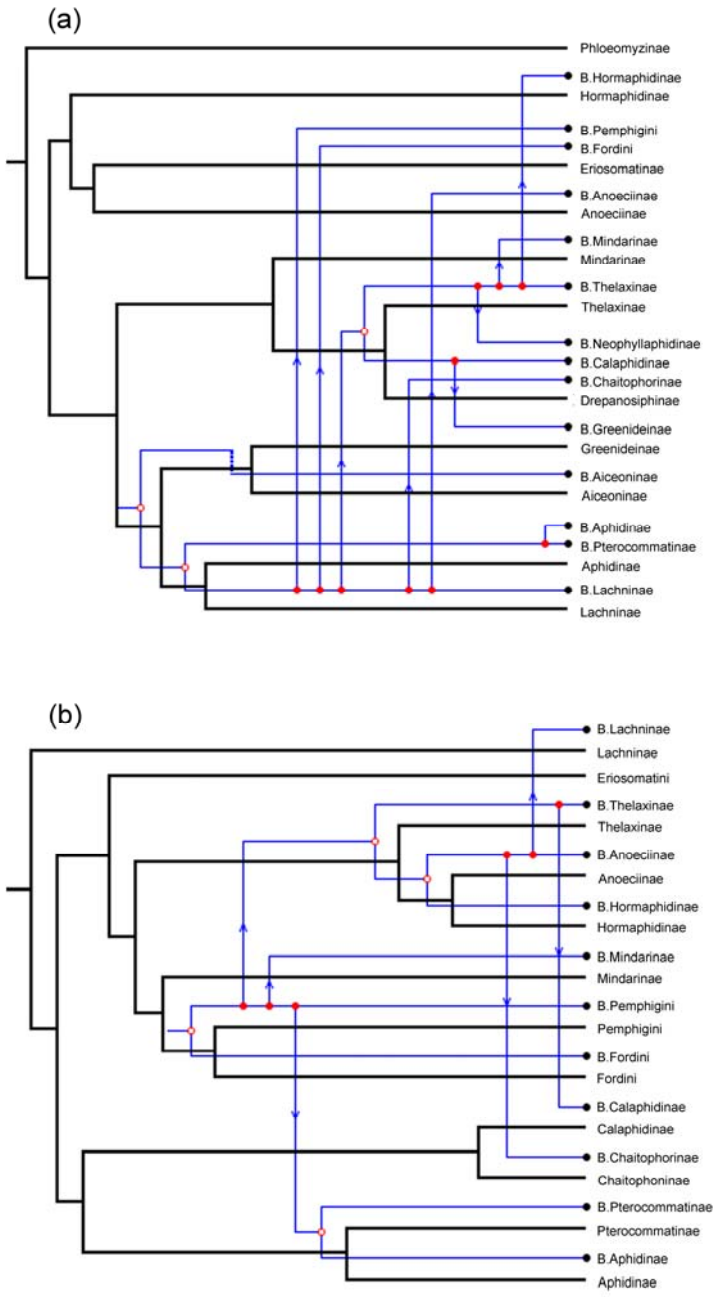

Figure S8 Cophylogeny of aphid and *Buchnera* from Jane. (a) The reconciled trees based on the morphology-based aphid tree and the simplified *gnd*/16S rDNA *Buchnera* tree. (b) The reconciled trees based on the simplified molecular-based aphid tree and the simplified *gnd*/16S rDNA *Buchnera* tree. Blue and black lines indicate the phylogenies of the *Buchnera* and aphids, respectively. Hollow red circles indicate cospeciation events; solid red circles indicate duplications; arrows indicate host switch events; dotted lines indicate loss events.

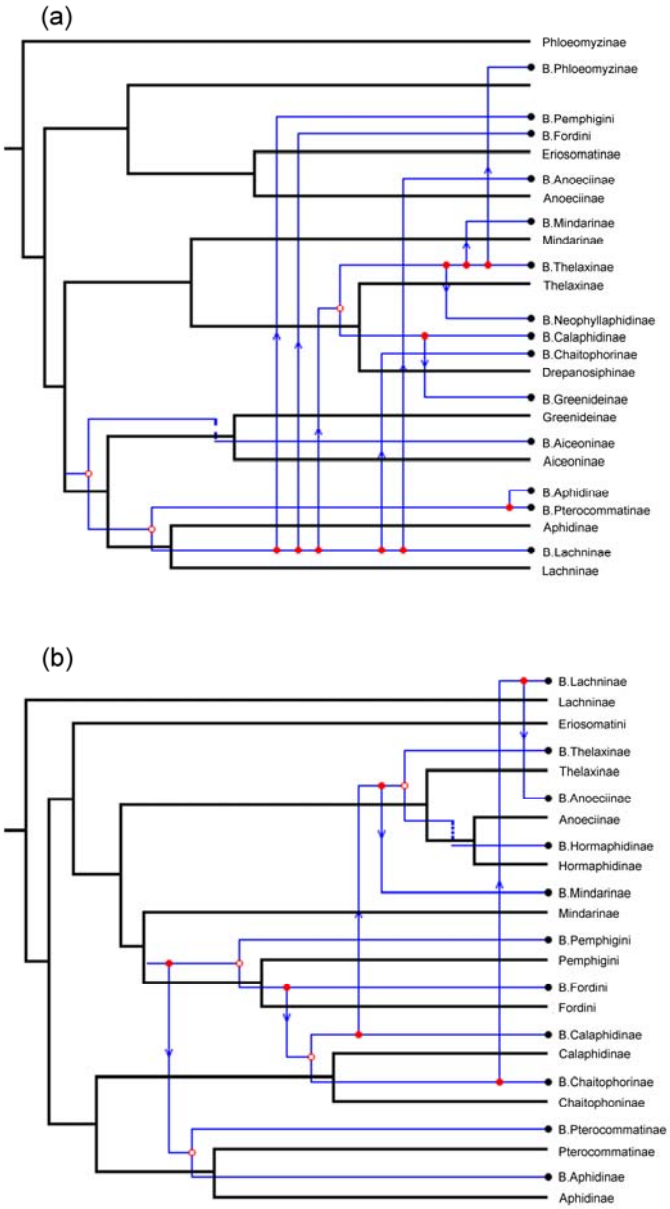

Figure S9 Cophylogeny of aphid and *Buchnera* from Jane. (a) The reconciled trees based on the morphology-based aphid tree and the simplified *gnd*/16S rDNA/*atpD Buchnera* tree. (b) The reconciled trees based on the simplified molecular-based aphid tree and the simplified *gnd*/16S rDNA/*atpD Buchnera* tree. Blue and black lines indicate the phylogenies of the *Buchnera* and aphids, respectively. Hollow red circles indicate cospeciation events; solid red circles indicate duplications; arrows indicate host switch events; dotted lines indicate loss events.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.346v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Table 1 Collection information for the aphid samples and GenBank accession numbers for *Buchnera* sequences used in this study

Subfamily	*Tribe*	*Aphid species*	*Locality*	*Date*	*Voucher number*	*GenBank accession number*		
						gnd	*atpD*	*16S rDNA*
Aiceoninae		*Aiceona himalaica*	Guizhou: Leigong Mountain	2005.6.4	16271	JQ920935	JX998132	JX998091
Aiceoninae		*Aiceona japonica*	Guizhou: Daozhen	2004.8.20	15616		JX998133	
Aiceoninae		*Aiceona osugii*	Fujian: Wuyi Mountain	2003.7.17	14517		JX998134	JX998092
Anoeciinae		*Anoecia fulviabdominalis*	Heilongjiang: Yichun	2005.8.09	17822	JQ920937		
Anoeciinae		*Anoecia fulviabdominalis*	Sinkiang: Urumchi	2002.8.16	13650		JX998136	JX998094
Anoeciinae		*Anoecia haupti*	Inner Mongolia: Yakushi	2005.8.17	17954		JX998137	
Aphidinae	Aphidini	*Aphis glycines*	Hunan: Changsha	2008.7.31	1994	JQ920938		JX998095
Aphidinae	Aphidini	*Cryptosiphum atriplicivorum*	Beijing: Lingshan Mountain	2007.8.14	20335	JQ920943	JX998143	JX998102
Aphidinae	Aphidini	*Rhopalosiphum maidis*	Shanxi: Taibai Mountain	2006.7.17	19378	JQ920959		JX998123
Aphidinae	Macrosiphini	*Brevicoryne brassicae*	Yunnan: Kunming	2006.4.22	18173	JQ920940	JX998138	JX998096
Aphidinae	Macrosiphini	*Cavariella salicicola*	Qinghai: Xining	2002.8.2	13589		JX998139	JX998097
Aphidinae	Macrosiphini	*Macrosiphoniella yomogifoliae*	Beijing: Lingshan Mountain	2007.8.14	20334	JQ920948	JX998148	JX998101
Aphidinae	Macrosiphini	*Metopolophium dirhodum*	Sichuan: Miyi	2005.4.20	17182	JQ920949	JX998149	JX998130
Aphidinae	Macrosiphini	*Myzus mushaensis*	Sichuan: Miyi	2005.4.20	17181	JQ920952	JX998151	JX998129
Calaphidinae	Calaphidini	*Betulaphis pelei*	Tibet: Maizhokunggar	2002.7.5	13548	JQ920939		
Calaphidinae	Calaphidini	*Chromaphis hirsutustibis*	Yunnan: Lijiang	2006.4.27	18225	JQ920942	JX998141	JX998099

Calaphidinae	Calaphidini	*Shivaphis celti*	Fujian: Wuyi Mountain	2003.7.17	14512	JQ920960		JX998125
Calaphidinae	Myzocallidini	*Tuberculatus yokoyamai*	Guizhou: Leigong Mountain	2005.6.5	16283			JX998131
Chaitophorinae	Atheroidini	*Sipha flava*	USA: Southern Illinois	2001.6.14	14181	JQ920961		
Chaitophorinae	Chaitophorini	*Periphyllus diacerivorus*	Beijing	2009.4.21	22198	JQ920954	JX998154	JX998117
Chaitophorinae	Chaitophorini	*Periphyllus koelrevteriae*	Beijing	2003.4.7	14216	JQ920955		
Eriosomatinae	Eriosomatini	*Eriosoma lanigerum*	Tibet: Nyingchi	2003.8.21	15412	JX998079		
Eriosomatinae	Eriosomatini	*Tetraneura sorini*	Beijing	2005.5.14	16188	JX998090	JX998159	JX998127
Eriosomatinae	Fordini	*Kaburagia rhusicola*	Yunnan: Yongsheng	2006.5.12	15699	JX998085	JX998146	JX998107
Eriosomatinae	Pemphigini	*Epipemphigus yunnanensis*	Yunnan: Lijiang	2006.4.27	18234		JX998144	JX998103
Eriosomatinae	Pemphigini	*Pemphigus tibetensis*	Tibet: Lhasa	2003.8.22	15429	JX998081	JX998153	JX998116
Eriosomatinae	Pemphigini	*Prociphilus liguatrifoliae*	Yunnan: Lijiang	2006.4.27	18235	JX998082	JX998155	JX998120
Greenideinae	Greenideini	*Greenidea cayratiae*	Fujian: Wuyi Mountain	2003.7.16	14499	JX998080		JX998105
Greenideinae	Greenideini	*Mollitrichosiphum montanum*	Tibet: Zayu	2005.8.22	18324	JQ920973		JX998114
Hormaphidinae	Cerataphidini	*Aleurodaphis blumeae*	Guizhou: Daozhen	2004.8.17	15597	JQ920936	JX998135	JX998093
Hormaphidinae	Cerataphidini	*Cervaphis quercus*	Guizhou: Daozhen	2004.8.24	15644	JQ920941	JX998140	JX998098
Hormaphidinae	Cerataphidini	*Ktenopteryx eosocallis*	Fujian: Wuyi Mountain	2003.7.4	14438	JQ920946		JX998108
Hormaphidinae	Hormaphidini	*Hormaphis similibetulae*	Tibet: Gongbo'gyamda	2002.8.2	13549	JQ920944		JX998106
Lachninae	Cinarini	*Cinara formosona*	Fujian: shanghang	2004.6.2	14882	JX998078	JX998142	JX998100
Lachninae	Cinarini	*Eulachnus pinitabulaeformis*	Shanxi: Foping	2005.5.14	17283	JX998084	JX998145	JX998104

Lachninae	Cinarini	*Schizolachnus orientalis*	Inner Mongolia: Chifeng	2005.8.21	17988	JX998083		JX998124
Lachninae	Lachnini	*Lachnus siniquercus*	Liaoning: Anshan	2005.7.18	17528	JX998088		JX998111
Lachninae	Lachnini	*Lachnus tropicalis*	Fujian: Wuyi Mountain	2003.7.18	14528	JX998086		JX998110
Lachninae	Lachnini	*Stomaphis sinisalicis*	Shandong: Qingdao	2002.10.14	13352	JX998087		JX998126
Lachninae	Lachnini	*Tuberolachnus salignus*	Jilin: Wufu	2004.8.7	14918	JX998089	JX998160	JX998128
Mindarinae		*Mindarus abietinus*	Shanxi: Taibai Mountain	2006.7.16	19364	JQ920950	JX998150	JX998112
Mindarinae		*Mindarus keteleerifoliae*	Yunnan: Kunming	2006.4.22	18171	JQ920951		JX998113
Neophyllaphidinae		*Neophyllaphis podocarpi*	Fujian: Wuyi Mountain	2003.7.4	14392	JQ920953	JX998152	JX998115
Phloeomyzinae		*Phloeomyzus passerinii*	Beijing	2003.6.5	14260	JQ920956		JX998119
Pterocommatinae		*Pterocomma kormion*	Jilin: Fusong	2004.8.7	14914	JQ920963	JX998156	JX998121
Pterocommatinae		*Pterocomma pilosum*	Heilongjiang: Mi shan	2005.8.2	17698	JQ920957	JX998157	JX998122
Pterocommatinae		*Pterocomma salicis*	Inner Mongolia: Yakeshi	2005.8.16	17933	JQ920958	JX998158	
Saltusaphidinae		*Izaphis sp.*	USA: Central Illinois	2002.8.26	14210	JQ920945		
Thelaxinae		*Kurisakia onigurumii*	Guizhou: Fanjing Mountain	2002.6.3	13303	JQ920947	JX998147	JX998109

Table 2 Primers and PCR conditions used for amplification of the *Buchnera* and aphid genes

Genes	*Primers (5'-3')*	*Annealing*	*Extension*	*Cycles*	*References*
16S rDNA	16sF: AGAGTTTGATCATGGCTCAGATTG 16sR: TACCTTGTTACGACTTCACCCCAG	65℃, 0.5min	72℃, 2min	35	Van Ham *et al.* (1997)
gnd	BamHI: CGCGGATCCGGWCCWWSWATWATGCCWGGWGG ApaI: CGCGGGCCCGTATGWGCWCCAAAATAATCWCKTTGWGCTTG	55℃, 0.5min	72℃, 1min	30	Clark *et al.* (1999)
atpD	ATPF: CGGGGATCCTGCAGTTTGGWGGWGCWGGWGTWGGWA AAAC ATPR: CGGGGATCCGTCGACGCATCWARATGWGCAAAWGTWGTWGCWGG	54℃, 0.5min	72℃, 1min	35	Martìnez-Torres *et al.* (2001)

Table 3 Characters of the sequence alignments for each gene and the combined data sets

Genes	*Number of ingroups*	*Length of alignments*	*Variable sites*	*Informative sites*	*Average p-distance*	*Nucleotide composition (T:C:A:G)*
16S rDNA	40	1344	378	270	0.070	21.7 : 21.3 : 28.6 : 28.4
gnd	41	903	710	631	0.373	37.1 : 11.1 : 37.2 : 14.6
gnd 1st and 2nd	41	602	419	351	0.279	31.0: 15.0: 35.5: 18.6
atpD	29	516	243	203	0.222	32.2 : 15.4 : 31.2 : 21.2
atpD 1st and 2nd	29	344	84	51	0.069	25.6: 20.1: 27.9: 26.4
[a]*gnd*/16S rDNA	35	1946	792	578	0.127	24.6 : 19.3: 30.7 : 25.4
[b]*gnd*/16S rDNA/*atpD*	22	2290	842	586	0.154	24.2 : 19.9 : 30.1 : 25.9

All the analyses were excluded outgroup.
a Exculding gnd third codon position.
b Excluding gnd and atpD third codon position

Table S1 Saturation pattern and evolutionary rates found for gnd and atpD genes

	Iss[a]	Iss.c[b]	p-value[c]
atpD 1st+2nd	0.1044	0.6827	0.0000**
atpD 3rd	1.0662	0.6754	0.0002**
gnd 1st+2nd	0.3180	0.7120	0.0000**
gnd 3rd	0.7300	0.6950	0.1965

[a] Index of substitution saturation (Iss) calculated using DAMBE software

[b] Iss.c stands for critical Iss values caculated for an asymmetrical tree.

[c] *p*-values indicate the probability that whether the score of Iss shows significant difference of Iss.c. *$0.01<p<0.05$, **$p<0.01$.

Table S2 Support value of clades to different datasets in the phylogenetic analyses performed in this study

	gnd			16S rDNA			B1 gnd/16S rDNA			B2 gnd/16S rDNA/atp		
	RAxml	PhyML	BI	RAxml	PhyML	BI	RAxml	PhyML	BI	RAxml	PhyML	BI
Aphidinae	83	74	1.00	n	n	n	100	100	1.00	92	86	1.00
Chaitophoninae	95	92	1.00	n	n	n	n	n	n	n	n	n
Calaphidinae	76	n	n	n	n	n	100	100	1.00	n	n	n
Greenideinae	63	90	0.99	98	93	1.00	100	96	1.00	n	n	n
Hormaphidinae	78	64	1.00	n	n	0.85	97	88	1.00	n	n	n
Lachninae	75	69	1.00	71	+	0.98	99	95	1.00	99	86	1.00
Mindarinae	100	100	1.00	99	100	1.00	100	100	1.00	n	n	n
Pemphigini	n	n	0.95	n	n	n	95	100	1.00	97	81	1.00
Pterocommatinae	100	100	1.00	98	100	1.00	100	100	1.00	100	100	1.00
Aphidinae + Pterocommatinae	95	93	1.00	62	n	1.00	100	97	1.00	100	100	1.00
Aphidinae + Pterocommatinae + Mindarinae	52	+	0.80	n	n	0.59	+	50	0.79	n	n	n
Chaitophoninae + Thelaxinae	+	50	0.99	n	n	n	n	n	n	n	n	n
Calaphidinae + Thelaxinae	n	n	n	n	n	n	+	+	+	n	n	n
Anoeciinae+ Lachninae	n	n	n	n	n	n	+	+	0.92	86	70	0.98

Node number refers to that in the working phylogenetic hypothesis (result of equally weighted parsimony analysis of total-evidence dataset). Support values are shown for the recovered clades. A dash is shown if MP bootstrap value is below 50%.

+, Clade recovered with ML bootstrap < 50%, or with posterior probabilities (PP) < 0.70.

n, Clade not recovered or not enough information available.

Table S3 Results of TreeMap.

Buchnera tree	Aphid tree	Mothed	Cospeciation	Duplication	Host Switch	Sorting event	Best Recostruction
gnd	Morphological-based	Heuristic Reasch	7 (p>0.05)	9	0	33	—
		Exact Reasch	7 (p>0.05)	4-7	2-5	12-20	1
	Molecular- based	Heuristic Reasch	5 (p>0.05)	6	0	37	—
		Exact Reasch	5 (p>0.05)	0-5	1-6	7-36	22
B1	Morphological- based	Heuristic Reasch	6 (p>0.05)	7	1	32	—
gnd/16S rDNA		Exact Reasch	6 (p>0.05)	0-7	1-8	16-32	426
	Molecular- based	Heuristic Reasch	4 (p>0.05)	6	0	37	—
		Exact Reasch	5 (p>0.05)	0-1	4-5	3-4	2
B2	Morphological- based	Heuristic Reasch	3 (p>0.05)	10	0	39	—
gnd/16S rDNA/atpD		Exact Reasch	4 (p>0.05)	0-7	2-9	9-25	1476
	Molecular- based	Heuristic Reasch	5 (p>0.05)	5	0	30	—
		Exact Reasch	5 (p>0.05)	0-4	1-5	5-24	18

Table S4 Results of Jane

Buchnera tree	Aphid tree	Cospeciation	Duplication	Host Switch	Loss event	Cost
gnd	Morphological-aphid	8	24	10	1	36(p>0.05)
	Molecular-aphid	8	12	6	1	20 (p>0.05)
B1	Morphological-aphid	8	20	9	1	31 (p>0.05)
gnd/16S rDNA	Molecular-aphid	8	12	6	0	18 (p>0.05)
B2	Morphological-aphid	6	20	9	1	31 (p>0.05)
gnd/16S rDNA/atp	Molecular-aphid	8	12	6	1	20 (p>0.05)

Table S5 Results of AU and SH test test among Buchnera alternative topologies

Tree	Δ -ln L[a]	SH-test[b]	AU-test[b]
gnd	best	0.858	0.790
gnd/16S rDNA	24.99639	0.447	0.318
gnd/16S rDNA/atp	49.10038	0.235	0.201

[a] Δ -ln L=*2*((ln *L1*)−ln(*L0*)) measures the difference in the likelihood between datasets being allowed to have different topologies (ln *L*1) and the highest likelihood obtained when all datasets are constrained to have the same topology (*L*0)

[b] Significance levels were determined in the tests. The scores indicate the probability that the score of the ML tree for a given dataset is significantly higher than that for the alternative topologies. $^{*}0.01<p<0.05$, $^{**}p<0.01$.

Table S6 Results of AU and SH test test among Buchenra and aphid trees with alternative datasets

Dataset\Tree	morphological-based aphid tree			Molecular-based aphid tree		
	Δ -ln L	SH-test[a]	AU-test[a]	Δ -ln L	SH-test[a]	AU-test[a]
gnd	133.637	0.0001^{**}	0.0001^{**}	104.657	0.0001^{**}	0.0001^{**}
gnd/16S rDNA	103.409	0.011^{*}	0.009^{**}	67.020	0.028^{*}	0.017^{*}
gnd/16S rDNA/atp	202.459	0.0001^{**}	0.0001^{**}	142.685	0.0001^{**}	0.0001^{**}

[a] Significance levels were determined in the tests. The scores indicate the probability that the score of the ML tree for a given dataset is significantly higher than that for the alternative topologies. $^{*}0.01<p<0.05$, $^{**}p<0.01$.