

Macroevolutionary Patterns in the Aphidini Aphids (Hemiptera: Aphididae): Diversification, Host Association, and Biogeographic Origins

Hyojoong Kim¹, Seunghwan Lee^{2*}, Yikweon Jang^{1*}

¹ Division of EcoScience and Research Institute of EcoScience, Ewha Womans University, Seoul, Republic of Korea, ² Department of Agricultural Biotechnology, Research Institute for Agriculture and Life Science, Seoul National University, Seoul, Republic of Korea

Abstract

Background: Due to its biogeographic origins and rapid diversification, understanding the tribe Aphidini is key to understanding aphid evolution. Major questions about aphid evolution include origins of host alternation as well as age and patterns of diversification in relation to host plants. To address these questions, we reconstructed the phylogeny of the Aphidini which contains *Aphis*, the most diverse genus in the family. We used a combined dataset of one nuclear and four mitochondrial DNA regions. A molecular dating approach, calibrated with fossil records, was used to estimate divergence times of these taxa.

Principal Findings: Most generic divergences in Aphidini occurred in the Middle Tertiary, and species-level divergences occurred between the Middle and Late Tertiary. The ancestral state of host use for Aphidini was equivocal with respect to three states: monoecy on trees, heteroecy, and monoecy on grasses. The ancestral state of Rhopalosiphina likely included both heteroecy and monoecy, whereas that of Aphidina was most likely monoecy. The divergence times of aphid lineages at the generic or subgeneric levels are close to those of their primary hosts. The species-level divergences in aphids are consistent with the diversification of the secondary hosts, as a few examples suggest. The biogeographic origin of Aphidini as a whole was equivocal, but the major lineages within Aphidina likely separated into Nearctic, Western Palearctic, and Eastern Palearctic regions.

Conclusions: Most generic divergences in Aphidini occurred in the Middle Tertiary when primary hosts, mainly in the Rosaceae, were diverging, whereas species-level divergences were contemporaneous with diversification of the secondary hosts such as Poaceae in the Middle to Late Tertiary. Our results suggest that evolution of host alternation within Aphidini may have occurred during the Middle Tertiary (Oligocene) when the secondary hosts emerged.

Citation: Kim H, Lee S, Jang Y (2011) Macroevolutionary Patterns in the Aphidini Aphids (Hemiptera: Aphididae): Diversification, Host Association, and Biogeographic Origins. PLoS ONE 6(9): e24749. doi:10.1371/journal.pone.0024749

Editor: Corrie S. Moreau, Field Museum of Natural History, United States of America

Received: April 29, 2011; **Accepted:** August 16, 2011; **Published:** September 15, 2011

Copyright: © 2011 Kim et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The paper was supported by RP-Grant 2010 of Ewha Womans University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: seung@snu.ac.kr (SL); jangy@ewha.ac.kr (YJ)

Introduction

The biology of aphids features some characteristics unusual in the animal kingdom, namely: polyphenism, alternation of sexual and asexual reproduction, and host alternation [1,2,3]. Evolution of these unusual characteristics is thought to be related to aphids' intricate ecological associations and evolutionary co-diversification with their host plants [4,5]. Although there is ample evidence of co-diversification of insects and their host plants across various taxa [6,7,8,9,10], major macroevolutionary patterns of co-diversification between them including age, patterns of diversification, and biogeographic origins often remain unclear [10]. For example, Lopez-Vaamonde et al. [11] proposed three hypotheses of temporal relationship between plant and insect diversifications: cospeciation, fast colonization, and delayed colonization. The cospeciation hypothesis is basically synchronized coevolution between phytophagous insects and their host plants, leading to congruent phylogenies and no time lag in diversifications between

them [11,12]. In both of the delayed colonization scenarios, phytophagous insects do not coevolve but instead colonize host plants that have already diversified in both fast and delayed colonization hypotheses [7,13]. Depending on the magnitude of evolutionary innovations required for using newly-diversified plants as resources, colonization may be fast or delayed [7,11,13].

Aphids are phloem-feeding insects, capable of infesting more than 40 plant families worldwide [2,3,14]. Based on fossil evidence and phylogenies, the ancestral aphids are hypothesized to have lived on woody host plants and reproduced sexually throughout the season [4,15,16]. Early in their evolution, aphids established parthenogenesis for their reproduction, as is found in all extant aphid taxa [4,5]. Typically, aphids undergo a series of all-female parthenogenetic generations, followed by a single generation of sexual reproduction [5]. This is called cyclical parthenogenesis, or holocycle [5]. Some aphids exhibit anholocycle in which the sexual generation is eliminated entirely; it is hypothesized that

anholocycly originated from holocycly based on loss of the sexual phase [5].

Another unusual feature of aphid evolution is the life cycle in relation to host plant use [1,5]. Monoecious aphids use the same type of host plants throughout their entire life cycles, whereas heteroecious aphids display host alternation between two distantly-related host plants, typically with the primary woody plants for sexual reproduction and the secondary herbaceous hosts for the parthenogenetic segment of a life cycle [4,5]. Therefore, all heteroecious aphids are holocyclic. There are in general three types of life cycle in extant aphids: (1) monoecy on trees, (2) heteroecy, and (3) monoecy on grasses [4,5]. Monoecy on trees is assumed to be the ancestral state for the family. Heteroecy is a more recently evolved state, in which a secondary host is acquired and the generations alternate host plants. Monoecy on grasses is then thought to have been derived through loss of the primary host tree species [4,5]. Less than 15% of aphids in the family Aphididae exhibit host alternation [4,5,17]. Heteroecy is most likely to have evolved in the Tertiary [4,5,16]. Contrary to the classical view of host alternation as a plesiomorphic trait inherited from a common Aphididae ancestor [18,19], Moran [4,20] suggested multiple gains within the subfamily Aphidinae. Later, based on a molecular phylogeny of Aphidinae, von Dohlen et al. [21] suggested that host alternation evolved twice: arising independently in both the tribes Aphidini and Macrosiphini. However, it still remains unclear when and how the different origins of host alternation arose for these groups, as their divergence times have never been estimated by a firm phylogenetic framework or compared with those of their host plants.

Approximately 5,000 described species of aphids belong to the family Aphididae (Hemiptera) [17], which may have diverged from the common ancestor of Adelgidae and Phylloxeridae in the Cretaceous [16,22]. Aphididae is divided into 27 subfamilies based on phenotypic, life cycle-specific, and host-specific variations [3,17]. Of the subfamilies, Aphidinae, which includes numerous agricultural pests, is the most diverse in the temperate regions of the Northern Hemisphere and subtropical regions [17,23]. Most modern taxa of Aphidinae likely diversified during the Tertiary [15,16]. Based on fossil records, at least 50% of the extant species of Aphidinae may have originated in the Middle to Late Tertiary [15,19]. The tribes Aphidini and Macrosiphini constitute Aphidinae, which has a sister relationship with the relatively small subfamily, Pterocommatinae. The tribe Aphidini contains more than 800 valid species, these aphids are relatively small and morphologically simple [14,17,24]. In a proposed alternative classification, Aphidini has been suggested as primitive to Macrosiphini, if Pterocommatinae and Macrosiphini form a clade [21]. In addition, Aphidini is considered to be a possible origin of Aphidinae, because this tribe is the only group that contains species indigenous to the Southern Hemisphere [21,25]. Aphidini is subdivided into two monophyletic subtribes, Aphidina and Rhopalosiphina [26]. The subtribe Aphidina contains the most species-rich genus, *Aphis*, whose rapid diversification may exemplify the evolutionary patterns of extant aphids [27]. Therefore, knowledge of taxon ages and patterns of diversification in Aphidini are critical to our understanding of aphid evolution [21].

We reconstructed the phylogeny of the tribe Aphidini and close relatives using DNA sequence data from one nuclear and four mitochondrial genes. Furthermore, we estimated divergence times using a molecular dating approach. Information generated in this study will be critical for understanding ages and patterns of diversification, origins of host alternation [16,21], and biogeographic origins in the aphids [14].

Methods

Ethical treatment of animals

Ethical approval was not required for work with the aphids, the subjects in this study, because aphids are invertebrates, and they are not listed as endangered species. Aphids are abundant almost everywhere in their natural ranges.

Taxon sampling and outgroup selection

A total of 80 ingroup species (59 Aphidina, 12 Rhopalosiphina, seven Macrosiphini, and two Pterocommatinae spp.) and seven outgroup species (two Hormaphidinae, one Lachninae, two Eriosomatinae, one Adelgidae, and one Phylloxeridae spp.) were used in this study (Table S1). We collected 46 species samples in the central and southern regions of the Korean Peninsula between 2003 and 2007, and, when available, used some sequences from previous studies [26,27,28]. DNA sequences of the ingroup species in Nearctic, European, and Australasian regions were obtained from GenBank (Table S1). The rest of the Aphidini sequences used in this study were derived from von Dohlen & Teulon [25], Turcinaviciene et al. [29], and Coeur d'acier et al. [30], to ensure representation of phylogenetically important taxa in each region. The sequences of the outgroup species, *Adelges cooleyi*, *Phylloxera* sp., *Hamamelistes spinosus*, *Melaphis rhois*, and *Schlechtendalia chinensis*, were also obtained from GenBank to get calibration points for dating analysis (Table S1, Figure S1) [16,21,22,23,29,31,32,33].

Within Aphidina, most species were sampled from the genus *Aphis*, which consists of four main species-groups (*craccivora*, *fabae*, *gossypii*, and *spiraeicola*), as well as from three other major subgenera (*Bursaphis*, *Protaphis*, *Toxoptera*). Two undescribed heteroecious species, *Aphis* sp.1 and sp.2 ex *Rhamnus* were included [28]. Two different types (type 1 and 2) of *A. gossypii* were collected from *Rhamnus*, its primary host; these were genetically different from other secondary host associated types, which were also included [28]. *Toxoptera aurantii* was included as a representative taxon characterized by a complete anholocyclic life [34]. Four major genera, *Hyalopterus*, *Melanaphis*, *Rhopalosiphum*, and *Schizaphis*, were included within Rhopalosiphina. *Aphis cottieri* Carver, *A. healyi* Cottier, *Casimira* sp., *Euschizaphis* sp.1, *Euschizaphis* sp.2, *Paradoxaphis aristoteliae* Sunde, and *P. plagianthi* Eastop, indigenous to the Southern Hemisphere, were included in order to determine whether Aphidinae or Aphidini originated there [25]. One sister clade of Aphidini, Macrosiphini, was represented by six genera (*Acyrtosiphon*, *Brevicoryne*, *Cryptosiphum*, *Lipaphis*, *Megoura*, and *Myzus*), which acted as representatives of two monophyletic lineages, Dactynotines and Myzines. For the other sister clade of Aphidini, *Pterocomma*+*Cavariella* were selected for construction of the expected clade of Pterocommatinae+*Cavariella* (P-C group), which had emerged as a monophyletic group in a previous phylogeny [21]. Two outgroups were selected at different taxonomic levels in order to set the calibration points precisely as well as to obtain reliable diversification times of Aphidinae corresponding to previous phylogenetic studies [16,22,35]. The first outgroup for fixing the calibration point diverging from the family Aphididae was the clade of Adelgidae+Phylloxeridae (*Adelges cooleyi* (Gillette) and *Phylloxera* sp.). The second outgroup for constraining the divergence point of the Aphididae crown clade consisted of three relative or distant subfamilies, Lachninae (*Cinara longipennis* (Matsumura)), Hormaphidinae (*Hamamelistes spinosus* Shimer and *Nipponaphis coreanus* (Paik)), Eriosomatinae (*Melaphis rhois* (Fitch), and *Schlechtendalia chinensis* (Bell)). Due to the rapid diversification of Aphididae subfamilies during the Cretaceous [16], it is still uncertain which group within Aphididae is the most basal lineage. Ortiz-Rivas and Martinez-Torres [36] recently

reported that Lachninae is the most basal lineage within Aphididae, but uncertainty remains due to sampling bias and constrained nodes. In contrast, Heie [15,19] suggested that Hormaphidinae and Eriosomatinae have more plesiomorphic morphological characters (e.g., shapes of antenna, secondary rhinaria, abdomen, and wing venation) than Lachninae. Therefore, three different subfamilies (Hormaphidinae, Lachninae, and Eriosomatinae) were used for calibrating the age of the Aphididae, in order to avoid uncertainties in the current phylogeny (von Dohlen and Moran, 2000; Ortiz-Rivas and Martínez-Torres, 2009). Two eriosomatids whose fossil and host plant data are available for divergence time calculation [37,38] were also used as a calibration point for the molecular dating analysis.

DNA sequencing and alignment

Total genomic DNA was extracted from single individuals using a DNeasy® Blood & Tissue Kit (QIAGEN, Inc., Düsseldorf) following the manufacturer's protocol. The primers for PCR amplification are listed in Table S2. LCO1490f and HCO2198 [39] were used to amplify partial cytochrome *c* oxidase I (COI). Primers 2993+ [40] and A3772 [41] were used to amplify partial tRNA-leucine+cytochrome *c* oxidase II (tRNA/COII). Primer F18 coupled with R18 [42] or CB2 [43] were used to amplify cytochrome *b* (CytB). Primers 12Sai [44] and 1473 [16] were used to amplify partial 12S rRNA+tRNA-valine+16S rRNA (12S/16S). Three primers, 12Sfr (a reverse of 12Sfi [45]), 1470a, and 1472 [16], were used as internal primers for sequencing. Primer EF3 coupled with EF2 [46] or EF6 [47] was used to amplify elongation factor 1 alpha EF1 α .

DNA fragments were amplified using AccuPower® PCR PreMix (BIONEER, Corp., Daejeon) in 20 μ l reaction mixtures containing 0.4 μ M of each primer, 20 μ M of dNTPs, 20 μ M of MgCl₂, and 0.05 μ g of genomic DNA template. PCR was performed using a GS482 thermo-cycler (Gene Technologies, Ltd., Essex) according to the following procedure: initial denaturation at 95°C for 5 min, followed by 34 cycles at 95°C for 30 sec; annealing temperature (43–45°C depending on the primer sets) for 30–50 sec; extension at 72°C for 30–60 sec, and final extension at 72°C for 5 min. The primer-specific annealing temperatures of each primer set were 43°C for COI, 42–45°C for tRNA/COII, 43–47°C for CytB, 48.5°C for 12S/16S, and 53–58°C for EF1 α . PCR products were visualized by electrophoresis on a 1.5% agarose gel. A single band was observed, purified using a QIAquick® PCR purification kit (QIAGEN, Inc.), and then sequenced directly using an automated sequencer (ABI Prism® 3730 XL DNA Analyzer). The sequences generated in this study were all deposited in GenBank (Table S1).

Raw sequences were examined and corrected using SeqMan™II (version 7.1.0, 2006; DNASTAR™). All DNA sequences for each fragment were aligned using Clustal X version 2.0.11 ([48]; with default settings). The intron splicing junctions of nuclear EF1 α sequences were identified and removed using MEGA 4.0 [49]. Ambiguous sites in 12S/16S containing the most gaps were removed using GBLOCKS 0.91b ([50]; default settings except for the allowed gap option where 'with half' was used). Uncorrected *P*-distances, number of substitutions, Transition (Ti)/Transversion (Tv) ratio, and nucleotide compositions for COI, tRNA/COII, CytB, and EF1 α were also obtained using MEGA.

Phylogenetic analysis

Maximum parsimony (MP) analyses were performed with PAUP*4.0b10 [51] using a heuristic search procedure with 1000 random additions of sequences and 10 trees held at each pseudoreplicate by following the TBR branch swapping method.

All characters were treated as unordered and equally weighted for MP analysis. Bootstrapping was conducted using 1000 replicates under the heuristic search procedure with 10 random-addition sequences. A partition-homogeneity test [52], as implemented in PAUP*, was performed using a heuristic search with 1000 replicates for significant phylogenetic analysis of the four mtDNA regions and EF1 α in two ways: i) individual mtDNA region vs EF1 α , ii) combined mtDNA dataset vs EF1 α . Taxa missing data for any dataset were automatically removed from the test.

For Maximum likelihood (ML) analysis, MrModeltest 2.0 [53], a simplified version of Modeltest [54,55,56], was used to select the best-fitting nucleotide substitution model, after which PAUP* settings were optimized based on the data of the selected model before searching. Then, ML analyses were performed under a partitioned scheme using RAxML 7.0.3 [57] with independent GTR+I+ Γ substitution models defined for each partition. The data were correspondingly partitioned into COI, tRNA/COII, CytB, 12S/16S, and EF1 α . Bootstrap analysis was also performed in RAxML, with 1000 bootstrap replicates from which a majority rule consensus tree was constructed in PAUP* for identification of supported clades.

Bayesian inference (BI) analyses were performed using MrBayes version 3.1.2 [58]. The best-fitting nucleotide substitution models (GTR+I+ Γ) and estimated parameters for each of the five partitions were selected using the hierarchical likelihood ratio test implemented in MrModeltest. Markov-Chain Monte Carlo (MCMC) analysis was carried out with one cold and three heated chains (temperature set to 0.1; starting from a random tree). The number of generations of the MCMC analysis and the tree sampling frequency were 10 million and 100 generations, respectively. The critical value for the topological convergence diagnostic of the preliminary tests was checked with MCMC options of 'stoprule = yes' and 'stopval = 0.01'. The burn-in parameter was estimated empirically by plotting $-\ln L$ against the number of generations using Tracer version 1.5 [59], and the trees corresponding to the first 20% generations were discarded. To ensure that the analyses were not trapped in local optima, five independent MrBayes runs were performed, after which topologies and posterior probabilities (PP) from different runs were compared for congruence purposes. We summarized the consensus tree using the post burn-in trees from all five runs in MrBayes (Fig. 1).

The effects of missing data or genes were assessed because 39 taxa among a total of 87 in this study were missing 15–68% of their sequences (Table S1). Generally, if enough characters have been sampled accurately to place all incomplete taxa on the tree, then the missing data will have little effect [60,61,62,63,64,65]. However, if a critical topological conflict or long branch attraction arises in the phylogenetic analyses due to inclusion of the taxa missing data, then the taxa cannot be used for estimation of divergence times [63,65,66,67]. To verify this, three different combined datasets were analyzed: the first one was a perfect concatenated dataset (48 taxa), the second one included taxa with at least three gene fragments (63 taxa [48 complete plus 15 missing 15–53% of their data]), and the third including all available taxa (87 taxa [63 previously described plus 24 missing 67–68% of their data]). MP, ML, and BI analyses were conducted following the same methodology, after which the nodal support values of significant group clusters (e.g., subfamily, tribe, subtribe, species-group) were compared for estimation of divergence times among the analyses of the three datasets.

Significant differences between topologies resulting from the above phylogenetic analyses, as well as topologies consistent with alternative hypotheses, were tested using the likelihood-based

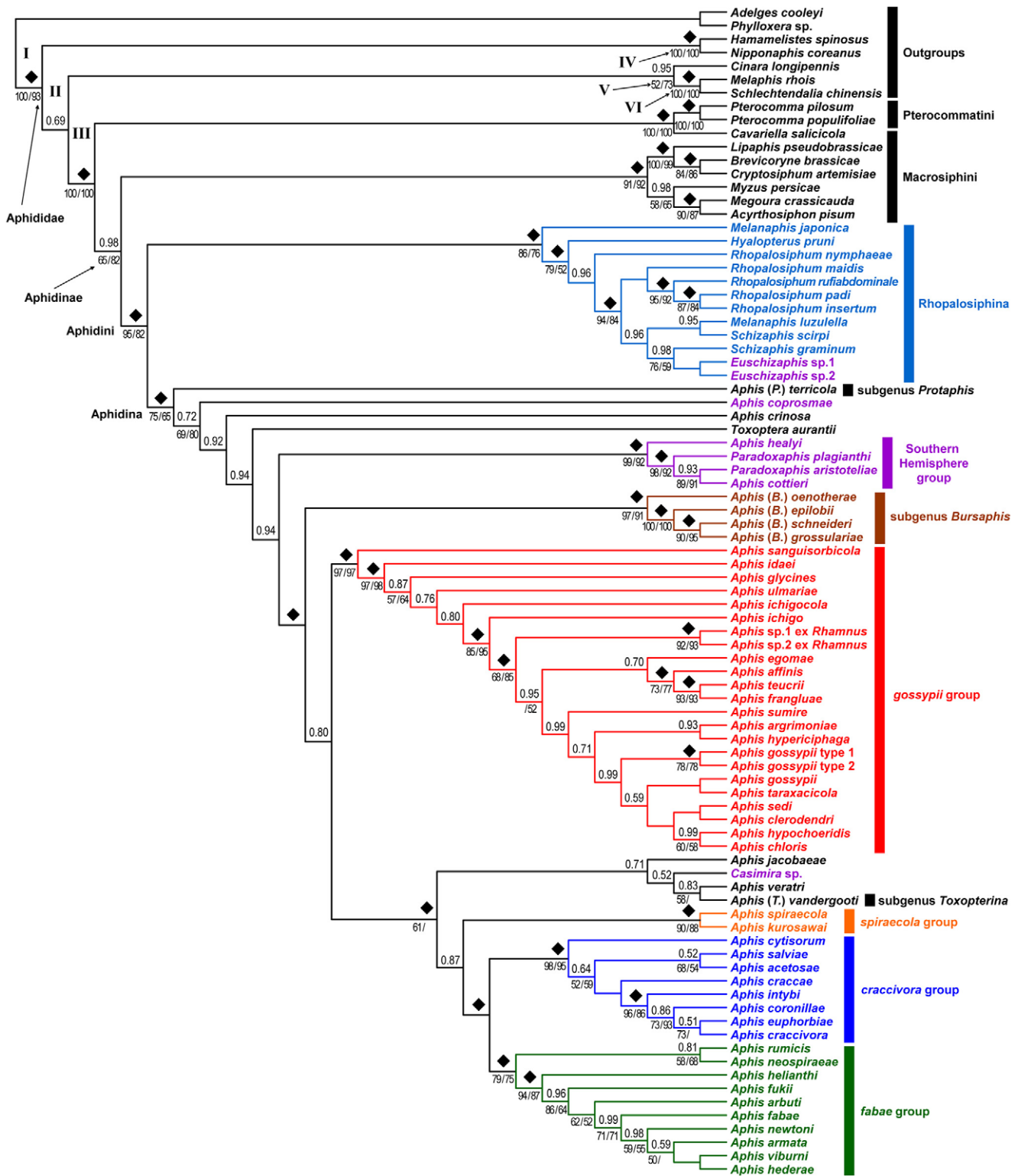


Figure 1. Cladogram representing the best ML topology tree of the Aphidini, Macrosiphini, and Pterocommatinae. Numbers above nodes indicate Bayesian posterior probabilities (PP), and numbers below nodes indicate ML bootstrap support values, followed by MP bootstrap support values. All support values are shown, if greater than 50%. ♦ indicates PP = 1.00. doi:10.1371/journal.pone.0024749.g001

Kishino-Hasegawa (KH) test [68] and an approximately unbiased (AU) test [69]. To perform the KH and AU tests, the first step was to reconstruct alternative tree topologies (fully-resolved) consistent

with the selected hypotheses using Mesquite version 2.6 [70]. ML heuristic searches using a GTR+I+Γ model for each partition that incorporated a topological constraint were conducted by RAxML

in order to produce the highest-likelihood topology that satisfied a given hypothesis. Second, PAML version 4.2b [71] was used to produce a log file (.lnf) for the log likelihoods of site-patterns of alternative trees given the concatenated dataset. The log file generated was submitted to CONSEL version 0.1i [72] to calculate the *P*-value for each alternative topology by the AU and KH tests.

Molecular dating and calibration points

Fossil records of aphids are restricted to the Late Cretaceous to the Tertiary, and most aphid fossils have been recovered from Canadian amber dated to 75–80 million years ago (MYA) or Baltic amber dated to 35–45 MYA [15,16,73]. Fossils of most extant subfamilies are known from the Eocene, but only two extant groups, Aphidinae and Neophyllaphidinae, are known from the Late Cretaceous [15,16]. Although there are few fossils of extant aphids that can be used to infer the exact time for molecular calibration, molecular dating for aphids was attempted in previous phylogenetic studies [16,38]. As the first reasonable estimation, von Dohlen and Moran [16] suggested divergence times of representative subfamilies in Aphididae based on analysis of the partial 12S and 16S rRNA genes. This estimate is based on crucial evidence from earlier research [37,38] that places the biogeographic isolation and divergence of the two sumac galling aphids, *Melaphis rhois* and *Schlechtendalia chinensis*, at 48–70 MYA. Moran et al. [38] previously estimated the age of the common ancestor of Aphididae to be 160–280 MYA based on the 16S rRNA sequences of the bacterial endosymbiont *Buchnera*, although later it was recalculated to be 84–99 MYA based on the common ancestor of these two melaphidines [16,37]. In addition, it was suggested that Aphidini and Macrosiphini diverged from one another at least 50 MYA based on fossil evidence (ca. 50 MYA) and Baltic amber. Moreover, their approximate divergence was inferred to have occurred between 50–70 MYA prior based on sequence divergences of aphid endosymbiotic *Buchnera* [74]. Moran [4,5] also suggested that aphids acquired host alternation ability between about 30–50 MYA based on fossil evidence. Recently, the divergence of Aphididae from two sister groups, Phylloxeridae and Adegidae, was inferred to have occurred between 120–150 MYA based on fossil evidence [22]. It seems valid for a molecular time estimation of Adelgidae [22], but most of the calibration points used in this estimate were obtained from earlier dating results of aphid subfamilies [16].

Therefore, based on previous studies that estimated divergence times [16,22], calibration points required for the molecular dating analyses were assigned as follows: i) the Aphidoidea crown clade (node I in Figure 1) was fixed at 150 MYA; ii) the Aphididae crown clade was constrained at a minimum age of 80 MYA and a maximum age of 100 MYA. However, since two nodes appeared in the phylogenetic analyses (nodes II and III in Figure 1), the same age constraint was applied for both nodes; iii) the divergence point of *M. rhois* and *S. chinensis* (node V in Figure 1) was constrained at a minimum age of 48 MYA and a maximum age of 70 MYA; iv) the divergence point of Aphidini and Macrosiphini (node 2 in Figure 2) was constrained at a minimum age of 50 MYA (Appendix 2). To reduce the uncertainties of the time estimation, two Bayesian inference-based programs, MULTIDIVTIME version 09.25.03 [75,76] and BEAST version 1.5.3 [77], were used to perform the molecular dating analyses. The geological time scale referenced is that of Gradstein and Ogg [78].

MULTIDIVTIME analysis

PAML/MULTIDIVTIME were used following the method of Rutschmann [79]. Although some taxa were missing from the

individual gene datasets, except for tRNA/COII (see Table S1), two package programs, ESTBRANCHES and MULTIDIVTIME, were able to account for the missing taxa [76]. To estimate the divergence times, a fully resolved topology of the combined dataset was obtained using RAxML (Figure 1), and this was also the best likelihood topology based on the KH and AU tests (see Results). At first, the BASEML program of PAML [71] was used to analyze the total molecular sequence data and parameters of the substitution model using the F84 model [68,80] for each gene separately based on individually optimized topologies. PAML2MODELINF was run to convert the BASEML output to useable data for ESTBRANCHES, which was then used to estimate branch lengths and their associated variance-covariance matrix using each output file from previous analyses. In this instance, the fully resolved target tree including the missing taxa was used. The outgroups were then pruned from the tree. The mean of the prior distribution of time from the ingroup root to the tip (rtm) was set to 0.9, and its standard deviation (rtmsd) was set to 0.1, in which one time unit represents 100 million years. Following the program manual recommendations, additional priors specified were rtrate = 0.35; rtratesd = 0.35; brownmean = 1.1; brownstd = 1.1; and bigtime = 100.0. The four nodes were constrained as follows: the Aphididae crown clade (ingroup root) and the clade of [Lachninae+Eriosomatinae]+Aphidinae were equally constrained at 80–100 MYA (L = 0.8, U = 1.0); the divergence point of *M. rhois* and *S. chinensis* at 48–70 MYA (L = 0.48, U = 0.7); the divergence point of Aphidini and Macrosiphini at a minimum age of 50 MYA (L = 0.5). Even though the most basal node did not require an additional constraint in MULTIDIVTIME, the constraint was maintained in order to compare its estimated time with that from BEAST, which requires constraining the same node. Other settings were left unchanged. The MCMC algorithm completed 300,000 initial burn-in cycles before the state of the Markov chain was sampled. Thereafter, the Markov chain was sampled every 100th generation until a total of 30,000 samples were collected. To test whether or not the Markov chain was convergent, three independent replicates were carried out.

BEAST analysis

A second analysis was performed using the BEAST software package 1.5.3 [77], which is designed to estimate divergence times using a Bayesian MCMC approach. At first, the software tool BEAUTi 1.5.3 of the BEAST package [77] was employed to design the run-file for BEAST. The uncorrelated lognormal model was used to describe the relaxed-clock, whereas GTR+I+ Γ was used to describe the substitution model. A Yule prior was used on the tree to simulate the process of speciation. In the BEAST analyses, the Aphidoidea crown ([Adelgidae+Phylloxeridae]+Aphididae) was fixed at 150 MYA, and the three other nodes were constrained according to the settings of the previous MULTIDIVTIME analyses. A preliminary test of MCMC run with 10 million generations was first performed to optimize the scale factors of the priori function. The final MCMC chain was run twice for 100 million generations sampled every 1000th generation. A 10% burn-in was discarded from the beginning of each run, and all samples were examined using Tracer 1.5 [59] to verify convergence and an effective sample size exceeding at least 200 for all parameters estimated. TreeAnnotator 1.5.3 of BEAST package [77] was used to summarize the mean parameter estimates and 95% highest posterior densities (HPDs), and then FigTree 1.3.1 [81] was used to visualize the results, including the confidence intervals.

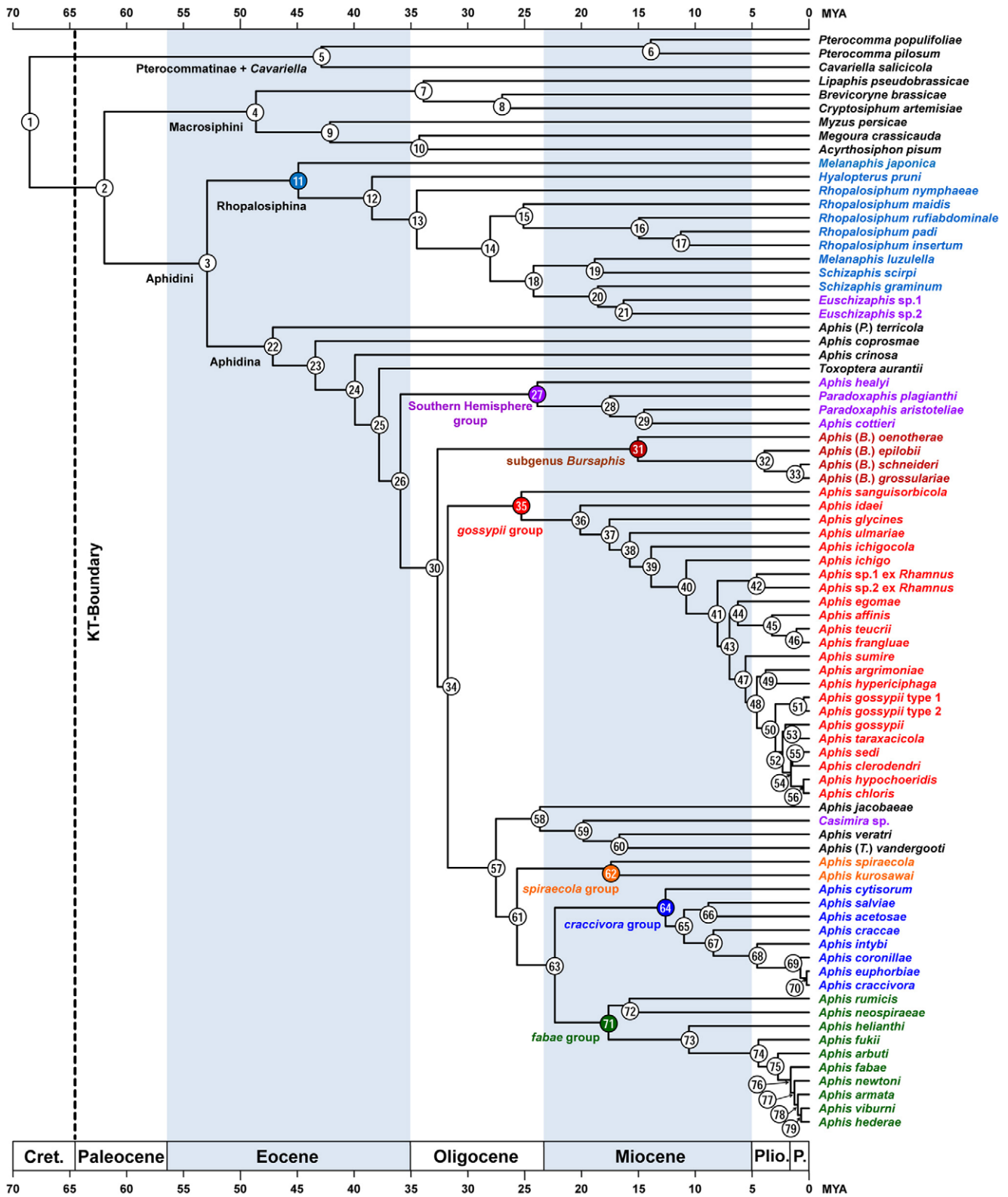


Figure 2. Chronogram showing the ages of origin and divergence times of the Aphidini, Macrosiphini, and Pterocommatinae. The topology corresponds to the best ML tree of Figure 1. The chronostratigraphic scale is given with absolute geological ages (MYA, million years ago; [78]). A node and species in the same color denote a clade. Numbers in circles refer to node numbers in Table 4 and Table S5. Cret. = Cretaceous. Plio. = Pliocene. P. = Pleistocene.

doi:10.1371/journal.pone.0024749.g002

Ancestral state reconstruction

Two ancestral states of the Aphidini, biogeography and host alternation, were reconstructed according to a Bayesian criterion [82] using BayesMultiState implemented in BayesTraits version 1.0 [83]. This method can allow for both polymorphism of character states and uncertainty in phylogeny. To reduce the uncertainty and arbitrary nature of choosing priors under MCMC, the reverse jump hyperprior approach (the rjhp command) was used as recommended [82,83]. For each test, combinations of hyperprior values (exponential or gamma, mean and variance) and rate parameter values were explored in order to find acceptance rates when running Markov chains between 20 and 40% (as recommended by [83]). A reverse jump hyperprior exponential (rjhp exp 0.0 30) distribution with a rate deviation prior of 10 was employed to analyze area, and a rjhp exp 0.0 2 with a rate deviation of 50 was used in the analysis of host alternation. Since tree branch length was important in this analysis, the 10 ML topology trees which showed similar best likelihood scores in the RAxML analyses were explored. The MCMC chain was run twice for 100 million generations sampled every 1000th generation after a burn-in of 10 million generations. The stationary phase during the MCMC run was observed by plotting the harmonic mean and then looking for a plateau, after which the means of each prior were calculated.

To reconstruct the ancestral state of host alternation, three states were identified: (0) monoecious holocyclic (mon. hol.) on an herbaceous plant, (1) mon. hol. on a shrubby or woody plant, (2) host alternation (heteroecy). Detailed information for coding the character state is given in Table S3. In this analysis, one anholocyclic species, *Toxoptera aurantii*, was regarded as holocyclic, whereas some species varying facultatively or genetically between anholocyclic and heteroecious (e.g., *Aphis gossypii*) were designated as host-alternating. Macrosiphini and Pterocommatinae were not included upon inferring two ancestral reconstructions for Aphidini.

For reconstruction of the ancestral state of area, the possible origin of the distribution of each species was coded into four regions based on the previous distribution records (e.g., Stroyan [84], Heie [85], Blackman and Eastop [14], Teulon and Stufkens [86], and Lee et al. [87]): (A) European (with some regions in the Western Palearctic), (B) Asian (with some regions in the Eastern Palearctic), (C) Australasian, and (D) Nearctic. Detailed information for coding the character states is given in Table S4. Because several taxa occurring in more than one region could not be coded to one state, the multiple character state option was used, which can be assigned in BayesMultistate: Palearctic (AB), Palearctic+Nearctic (ABD), Cosmopolitan (ABCD). According to the BayesTraits manual [83], the code AB signifies that a trait can be in states A or B (with equal probability) but not in states C or D. Tropical areas, i.e., Afrotropical, Indo-Malayan, and Neotropical regions, were excluded since aphids are thought to have originated in temperate regions, especially the Northern Hemisphere [21,25].

Results

Phylogenetic analysis

In the comparison of individual gene datasets, CytB had the largest proportion of informative characters (32.2%) as well as the greatest pairwise sequence divergence (8.0%) between ingroup species among the five DNA regions (Table 1). In contrast, the nuclear EF1 α had the smallest sequence divergence among all sequence regions, and the sequence divergence of 12S/16S was the lowest among all mitochondrial regions. Regarding the Ti/Tv ratio, three mitochondrial genes showed moderate ratios (ca. 1.25),

whereas 12S/16S showed predominance of Tv (0.389). On the contrary, EF1 α showed a predominance of Ti (2.333). The partition-homogeneity test [52] showed no significant evidence of phylogenetic conflicts between the two paired regions or within the combined dataset ($0.07 \leq P \leq 0.91$). Thus, these five regions are expected to account for different taxonomic levels, suitable for this phylogenetic reconstruction.

The effects of the missing data or genes (15–68%) were assessed using MP, ML, and BI analyses with the three different combined datasets (Table 2). Hereafter, the combined datasets (CDS) with 48, 63, and 87 taxa are abbreviated to CDS-48, CDS-63, and CDS-87, respectively. For each CDS, the best-fitting model of nucleotide substitution was GTR+I+ Γ in both the ML and BI analyses. No topological conflict was identified among the three datasets, as the most important nodes in each dataset were recovered. The statistical support values of the datasets were compared with the 17 important nodes responsible for the subfamily, tribal, subtribal, and species-group clusters. Bootstrap values estimated in the MP and ML analyses were significantly affected by inclusion of the taxa with missing data, whereas posterior probabilities of the BI analysis were relatively less sensitive. However, the ML or BI support values between CDS-63 and CDS-87 increased on several nodes upon inclusion of the taxa with missing data (Table 2). This implies that the taxa with missing data could be used to corroborate each clade without topological conflict. In the molecular dating analysis, the topologies in the BI analysis were chosen rather than those in the ML or MP analysis, since both BEAST and MULTIDIVTIME estimated the divergence times under the Bayesian algorithm-based clock model [76,77]. Because there seemed to be no significant difference in support values of the BI analysis among three datasets, CDS-87 was used for both phylogenetic reconstruction and estimation of divergence times.

The ML tree based on the best likelihood score corresponds to the 50% majority rule consensus tree of the BI analysis, except for some unresolved clades (Figure 1). Due to large genetic distances between the outgroup and ingroup species, the cladogram is illustrated showing only relationships instead of the phylogram. The relative genetic distances between ingroup species can be seen in Figures 3 and 4. The clade consisting of Aphidinae+Pterocommatinae (node 2 in Figure 2) was well supported in all analyses. In this study, the P-C group was the most basal tribe within Aphidinae, but it was not robustly supported with 0.98 PP or the 65% ML bootstrap value. Except for the P-C group, all other tribal and subtribal clades received 1.0 PP and a ML-bootstrap value ranging from 75 to 95%. The tribe Macrosiphini was separated from the tribe Aphidini, which in turn was subdivided into two monophyletic subtribes, Aphidina and Rhopalosiphina. Within Rhopalosiphina, *Melanaphis japonica* was sister to the remaining rhopalosiphine species with 1.0 PP and a 79% ML bootstrap value. In the BI analysis, *Melanaphis luzullella* was not clustered with *M. japonica* but was closely related with *Schizaphis* species. Within Aphidina, *Aphis terricola*, *A. coprosmae*, and *A. crinosa* appeared sequentially in the basal nodes. These three species are suggested to be the most basal taxa of all Aphidina species, even though the sister clade of *A. crinosa* received low support values (0.72–0.94 PP). Although these three species did not form a clade, *A. crinosa* and *A. coprosmae* were most likely transferred to the subgenus *Protaphis*, because their morphological characters were consistent with those of *Protaphis* [85,86,88]. As the sister group of the node of *Toxoptera aurantii*, four Southern Hemisphere species clustered as a sister group consisting of the remaining *Aphis* species, which were robustly supported in the BI analysis (1.0 PP). Except for the genus *Bursaphis*, most *Aphis* species

Table 1. Characteristics of DNA sequences and three combined datasets.

	Single individual datasets					Combined datasets		
	COI	tRNA/COII	CytB	12S/16S	EF1 α	CDS-48	CDS-63	CDS-87
Number of taxa	60	87	76	51	62	48	63	87
Aligned sequence length (bp)	658	702	737	1601	802	4500	4500	4500
Variable sites (%)	247 (37.5)	310 (44.2)	325 (44.1)	600 (37.4)	230 (28.7)	1568 (34.8)	1682 (37.4)	1712 (38.0)
Informative characters (%)	208 (31.6)	224 (31.9)	237 (32.2)	324 (20.2)	180 (22.4)	1046 (23.2)	1156 (25.7)	1173 (26.1)
Nucleotide composition (T:C:A:G)	41:14:35:10	39:12:41:8	43:13:35:9	46:5:38:11	26:22:28:24	40:12:36:12	40:12:36:12	40:12:36:12
Pairwise sequence divergence*	7.5 \pm 2.2	6.0 \pm 2.2	8.0 \pm 2.8	3.8 \pm 1.5	3.6 \pm 2.3	5.3 \pm 1.9	6.2 \pm 2.2	6.0 \pm 2.1
Ti/Tv ratio	1.216	1.281	1.250	0.389	2.333	0.969	1.135	1.250

To reconstruct the phylogeny of Aphidini, we used one nuclear and four mitochondrial DNA sequences. Due to missing sequence data, we compared three different combined datasets.

*indicates uncorrected *P*-distance (mean \pm S.D.) among ingroup species.

doi:10.1371/journal.pone.0024749.t001

were partitioned into two subclades, the *gossypii* group and the *craccivora*+*fabae*+*spiraecola* groups. Each of these four species groups was highly supported by 1.0 PP and a ML bootstrap value ranging from 79 to 98%.

The results of the KH and AU tests of the alternative tree topologies are summarized in Table 3. The two alternative sister relationships with Pterocommatinae were not significantly different, but their confidence values (0.086–0.171) were too low to replace the best topology. Of ten alternative tree topologies tested with respect to the basal position within Aphidina, seven were rejected (*P*<0.05). In particular, the basal location of all six Southern Hemisphere species within Aphidina was accepted even with low confidence values. For the tests of the basal position within Rhopalosiphina, only the alternative position of the genus *Hyalopterus* was allowed with narrow confidence values. Three

possible monophylies were tested, and then the monophyly of the genus *Rhopalosiphum* was rejected. Consequently, although eight alternative topologies were accepted (Table 3), they received much lower confidence values, ranging from 0.062 to 0.253, than did the best topology.

Divergence times

The estimated divergence times for the 33 selected nodes of the chronogram (Figure 2) are summarized in Table 4, and those for all nodes are shown in Table S5. Mean age estimates of the 33 nodes were slightly different, averaging 1.82 MYA between the MULTIDIVTIME and BEAST analyses. However, 95% HPDs of the BEAST analyses generally overlapped with those of the MULTIDIVTIME analyses, suggesting that the time estimates of the two programs were largely congruent. Based on the results of the

Table 2. Statistics of support values estimated from three combined datasets.

Node no.	Node description	87 taxa			63 taxa			48 taxa		
		BI	ML	MP	BI	ML	MP	BI	ML	MP
1	Aphidinae+Pterocommatinae	1.00	100	100	1.00	100	100	1.00	100	100
2	Macrosiphini+Aphidini	0.98	65	82	0.98	63	83	1.00	69	69
3	Aphidini	1.00	95	82	1.00	98	89	1.00	96	77
4	Macrosiphini	1.00	91	92	1.00	91	98	1.00	87	94
5	Pterocommatinae	1.00	100	100	1.00	100	100	1.00	100	100
11	Rhopalosiphina	1.00	86	76	1.00	84	80	1.00	82	65
12	clade sister to <i>M. japonica</i>	1.00	79	52	1.00	80	53	1.00	77	58
22	Aphidina	1.00	75	65	1.00	100	97	1.00	100	99
24	clade sister to <i>A. crinosa</i>	0.94	37	20	0.83	43	43	1.00	65	68
25	clade sister to <i>T. aurantii</i>	0.94	24	24	0.76	23	28	1.00	92	68
30	clade sister to Southern Hemisphere group	1.00	39	23	0.86	48	31	-	-	-
34	clade of four species groups with other spp.	0.80	29	15	0.86	35	18	1.00	41	-
35	<i>gossypii</i> group	1.00	97	97	1.00	100	99	1.00	100	100
61	<i>craccivora</i> + <i>fabae</i> + <i>spiraecola</i> groups	0.87	36	26	1.00	84	62	1.00	92	77
63	<i>craccivora</i> + <i>fabae</i> groups	1.00	42	35	1.00	50	55	0.51	43	48
64	<i>craccivora</i> group	1.00	98	95	1.00	100	100	-	-	-
71	<i>fabae</i> group	1.00	79	75	1.00	96	96	1.00	97	99

We used BI, ML, and MP analyses to compare the datasets. Node no. refers to nodes of phylogeny in Figure 2.

doi:10.1371/journal.pone.0024749.t002

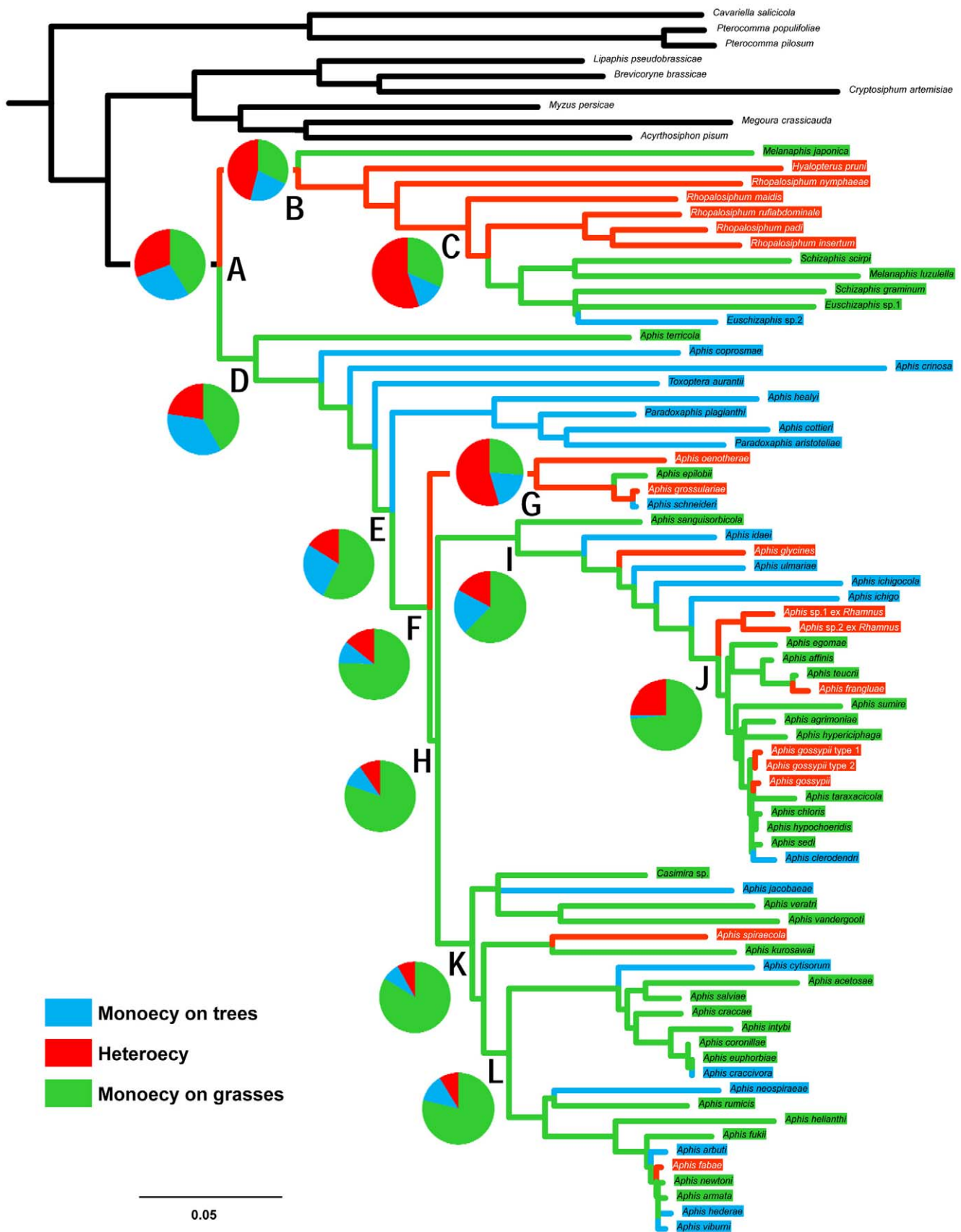


Figure 3. Ancestral state reconstruction for host alternation. The ancestral states are classified into monoecy on trees (blue), heteroecy (red), and monoecy on grasses (green). The topology is derived from the ML tree of Figure 1. Pie charts indicate the relative likelihoods at respective nodes (A–L). Terminal taxa and their respective branches are color-coded for state of host use. The scale is a nucleotide substitution rate of 0.05. doi:10.1371/journal.pone.0024749.g003

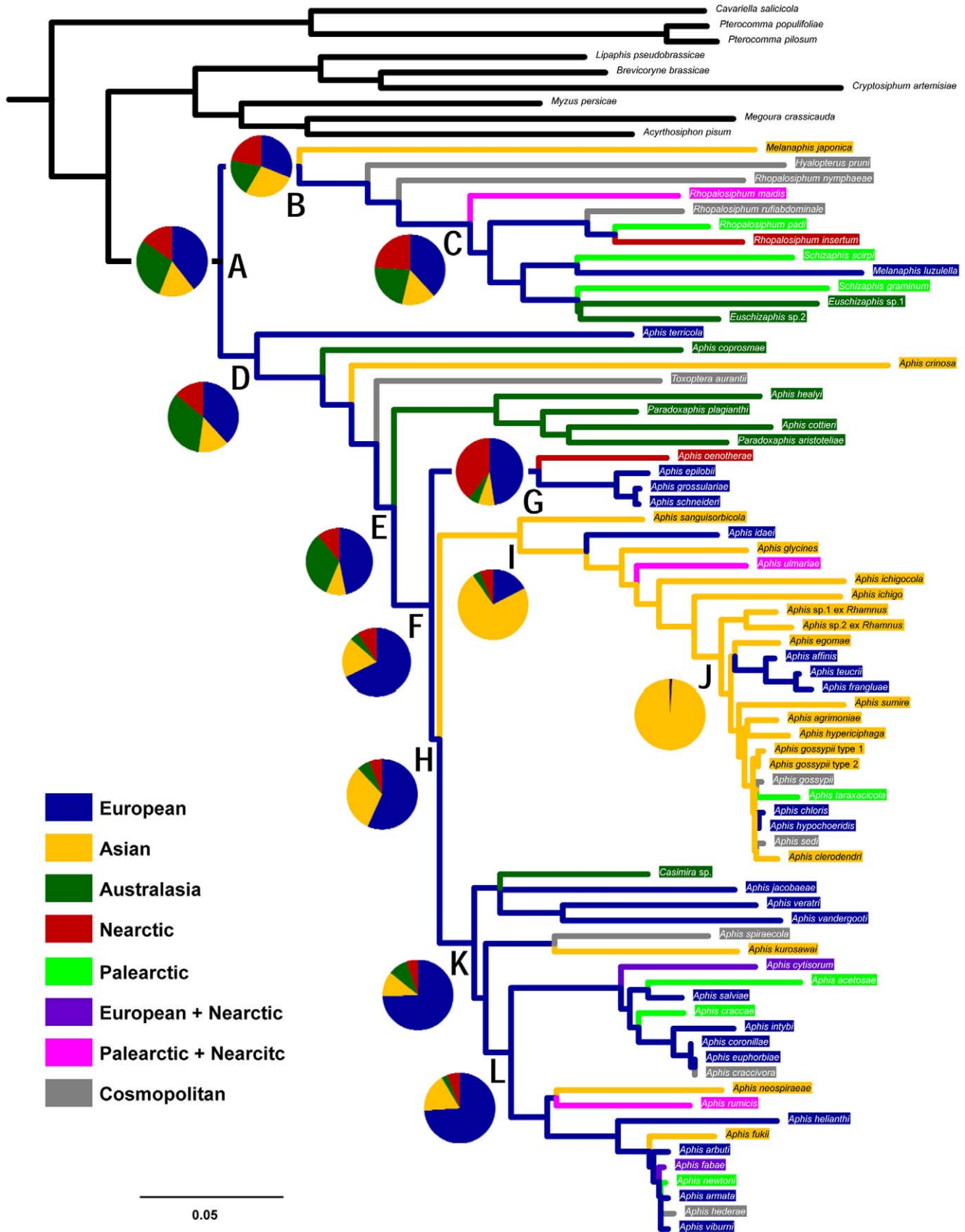


Figure 4. Ancestral state reconstruction for biogeographic origin. The topology is derived from the ML tree of Figure 1. Pie charts indicate the relative likelihoods at respective nodes (A–L). Terminal taxa and their respective branches are color-coded for state of host use. The scale is a nucleotide substitution rate of 0.05. Palearctic, European+Nearctic, Palearctic+Nearctic, and cosmopolitan states were coded as multistate and thus do not appear in pie charts.
 doi:10.1371/journal.pone.0024749.g004

Table 3. Comparison between the best (1) and the alternative (2–20) topologies.

Topology	Description of alternative topology	Rank	Obs	KH	AU
1	Best ML tree from RAxML	1	---	0.522	0.823
2	Secondly best ML tree from RAxML	2	0.3	0.478	0.709
3	(Pterocommatinae+Aphidini)+Macrosiphini	7	17.8	0.086	0.129
4	(Pterocommatinae+Macrosiphini)+Aphidini	4	14.7	0.089	0.171
5	Basal position of <i>Bursaphis</i> within Aphidina	16	46.3	0.019*	0.024*
6	Basal position of <i>gossypii</i> group within Aphidina	18	54.3	0.022*	0.021*
7	Basal position of (<i>craccivora</i> + <i>fabae</i> + <i>spiraecola</i> groups)+node 58 within Aphidina	15	44	0.019*	0.022*
8	Basal position of node 30 within Aphidina	13	31.8	0.023*	0.014*
9	Basal position of <i>craccivora</i> + <i>fabae</i> + <i>spiraecola</i> groups within Aphidina	17	52.9	0.01*	0.016*
10	Basal position of all Southern Hemisphere species (non-monophyly) within Aphidina	14	39.4	0.022*	0.019*
11	Basal position of <i>A. coprosmae</i> +Southern Hemisphere group within Aphidina	11	24	0.076	0.157
12	Basal position of <i>A. coprosmae</i> within Aphidina	3	11.2	0.09	0.253
13	Basal position of <i>A. crinosa</i> within Aphidina	6	17.6	0.025*	0.059
14	Basal position of <i>T. aurantii</i> within Aphidina	12	25.2	0.067	0.108
15	Monophyly of four species groups excluding node 58	8	22.3	0.062	0.092
16	Monophyly of <i>Rhopalosiphum</i>	10	23.8	0.031*	0.04*
17	Monophyly of <i>Melanaphis</i>	9	22.3	0.065	0.086
18	Basal position of <i>Hyalopterus</i> within Rhopalosiphina	5	15.5	0.081	0.104
19	Basal position of <i>Rhopalosiphum</i> (non-monophyly) within Rhopalosiphina	19	64.8	0.001*	<0.001*
20	Basal position of (<i>Schizaphis</i> + <i>Euschizaphis</i> [non-monophyly]) within Rhopalosiphina	20	65.2	0.007*	0.006*

Nodes 30 and 58 are those referred in Figure 2.

*signifies that the hypothesis received a *P* value<0.05 and can be rejected.

doi:10.1371/journal.pone.0024749.t003

MULTIDIVTIME and BEAST analyses, the divergence point of the P-C group and Aphidinae was estimated to be immediately before the K-T boundary (67–68 MYA), whereas the divergence of Aphidini and Macrosiphini occurred after that (62 MYA). The divergences within the tribal and subtribal clades arose in the Early to Middle Eocene (42–55 MYA). Within the Pterocommatinae+Cavariella group, the divergence between Pterocamma and Cavariella was dated to ca. 42 MYA. The divergences of the rhopalosiphine genera occurred over a considerable interval. That is, *Melanaphis* diverged first near the Middle Eocene (45–50 MYA), whereas *Schizaphis* emerged during the Late Oligocene (24–30 MYA). Within Aphidina, the divergence times of the most extant members in the subgenus or species-group were estimated to be in the Late Oligocene to Middle Miocene (12–25 MYA). Some morphologically cryptic species in *gossypii*- and *fabae*-groups arose mostly after the Pliocene (<5 MYA). In summary, most generic divergences in Aphidini occurred in the Middle Tertiary, and species-level divergences occurred in the Middle or Late Tertiary.

Evolution of host plant association and host alternation

The estimation of divergence times suggests that Pterocommatinae and Aphidinae likely diversified during the radiation period of their host plants (Tables 4, 5). The divergence between Aphidinae and the P-C group (node 1 in Figure 2) likely occurred along with early diversification of Rosaceae [89,90]. The divergence times of *Pterocamma* and *Cavariella* (node 5) were inferred to be near the earliest fossil record of Salicaceae [11] and Araliaceae [89]. The divergence between Aphidini and Macrosiphini (node 2) in the Middle Paleocene overlapped the periods suggested by the earliest fossil record of Rosaceae [91] and by the molecular dating results for Rosaceae [89]. In addition, four basal

divergences for Macrosiphini, Aphidini, Aphidina, and Rhopalosiphina (nodes 3, 4, 11, 22) within Aphidinae were embedded within the initial divergence periods of Rosaceae [89,91].

Three *Protaphis*-like species, *Aphis terricola*, *A. coprosmae*, and *A. crinosa*, placed basally within Aphidina, diverged during the Eocene and corresponded to the appearances of their host plants, Asteraceae, Rubiaceae, and Oleaceae, respectively [92,93,94]. Divergences of the *Melanaphis* (node 11) at 45–50 MYA and the *Hyalopterus* (node 12) at 38–43 MYA were also similar to the appearances of *Prunus* or *Spiraeoideae* [11,89]. The divergence point of the most recent common ancestor (MRCA) of the *gossypii* group was estimated at 20–25 MYA and overlapped the divergence times of its primary hosts, *Rhamnus* and *Frangula* [95], assuming that the MRCA of the *gossypii* group (node 35) associated with these hosts. Therefore, the divergence times between generic or subgeneric level taxa of aphids and their primary hosts are almost consistent. However, *Ribes*, a primary host genus of the subgenus *Bursaphis* that diverged in the Miocene, likely diversified in the Late Cretaceous [96].

In comparison of the divergence times between aphid taxa and their secondary hosts, one species of the *gossypii* group, *Aphis glycines* (node 37), was estimated to have diverged at 15–17 MYA, which was precisely nested within the estimated times of *Glycines*, its secondary host species [97,98]. The divergence times of *Epilobium* and *Oenothera* [99,100] are closer to those of *Bursaphis* species (nodes 31). However, the estimated divergence times of *Hyalopterus* differ considerably from those of *Phragmites* (17–20 MYA) [101], which is the sole secondary host genus of *Hyalopterus* [14]. Similarly, the divergence of *Melanaphis* (node 11) is much earlier than that of the host *Miscanthus* [101]. Therefore, the divergence times of secondary hosts are more consistent with those of aphid species than those of genera.

Table 4. Estimated divergence times for selected nodes.

Node no.	Node explain	BEAST		Multidivtime	
		Time	95% HDP range	Time	95% HDP range
II	Aphididae	97.0	(91.5–100.0)	94.2	(85.1–99.7)
III	(Lachninae+Eriosomatinae)+ingroup clade	94.2	(87.9–99.3)	92.4	(83.5–98.7)
IV	Hormaphidinae	65.7	(48.5–82.5)	70.8	(57.5–84.0)
V	Lachninae+Eriosomatinae	83.7	(73.9–93.4)	87.7	(77.3–96.4)
VI	Eriosomatinae	52.4	(48.0–59.6)	51.5	(48.1–59.4)
1	(Pterocommatinae+Cavariella)+Aphidinae	68.5	(58.8–78.5)	67.2	(58.7–77.4)
2	Macrosiphini+Aphidini	62.0	(52.9–71.0)	62.2	(54.5–71.9)
3	Aphidini	52.9	(45.0–61.4)	55.0	(50.2–63.9)
4	Macrosiphini	48.6	(38.4–58.9)	48.9	(39.9–59.3)
5	Pterocommatinae+Cavariella	42.9	(29.6–55.9)	42.2	(31.9–53.8)
11	Rhopalosiphina	44.9	(36.9–53.2)	50.6	(43.8–59.7)
12	divergence of <i>H. pruni</i>	38.4	(31.2–46.4)	42.0	(34.5–51.2)
14	<i>Rhopalosiphum</i> + <i>Schizaphis</i> clade	28.0	(22.2–34.6)	33.0	(25.7–41.8)
18	<i>Schizaphis</i> (including <i>M. luzulella</i>)	24.2	(18.5–30.4)	29.7	(22.3–38.3)
22	Aphidina	47.2	(38.6–55.6)	50.4	(42.6–60.0)
23	divergence of <i>A. coporosmae</i>	43.4	(36.0–51.1)	45.6	(38.4–54.6)
24	divergence of <i>A. crinosa</i>	39.9	(32.9–47.2)	43.1	(36.1–52.0)
25	divergence of <i>T. aurantii</i>	37.8	(31.4–44.6)	39.1	(32.0–47.8)
26	divergence of Southern Hemisphere group	35.9	(29.6–42.5)	35.7	(28.5–44.4)
27	clade of Southern Hemisphere group	23.9	(16.9–30.8)	24.8	(16.6–34.5)
30	clade sister to Southern Hemisphere group	32.7	(27.1–38.7)	30.4	(23.8–38.7)
31	clade of subgenus <i>Bursaphis</i>	15.0	(8.2–22.8)	13.6	(6.4–22.4)
34	clade of four species groups with other spp.	31.8	(26.3–37.6)	28.0	(21.7–35.9)
35	<i>gossypii</i> group	25.3	(20.2–30.6)	20.4	(14.7–27.6)
37	divergence of <i>A. glycines</i>	17.5	(13.5–21.8)	15.0	(10.5–20.9)
41	divergence of two <i>Aphis</i> spp. ex <i>Rhamnus</i>	8.0	(6.0–10.1)	6.4	(3.9–9.8)
48	morphologically cryptic species with <i>A. gossypii</i>	4.6	(3.1–6.1)	2.9	(1.5–4.9)
61	<i>craccivora</i> + <i>fabae</i> + <i>spiraecola</i> groups	25.7	(20.8–30.8)	23.1	(17.3–30.3)
62	<i>spiraecola</i> group	17.4	(11.4–23.3)	17.1	(11.4–23.9)
63	<i>craccivora</i> + <i>fabae</i> groups	22.4	(17.6–27.2)	19.8	(14.3–26.7)
64	<i>craccivora</i> group	12.6	(8.7–16.9)	11.7	(6.2–19.0)
71	<i>fabae</i> group	17.6	(13.2–22.3)	16.4	(11.6–22.4)
74	morphologically cryptic species with <i>A. fabae</i>	4.5	(2.5–6.6)	3.6	(1.8–6.1)

Node numbers refer to those in Figures 1 and 2. Roman numerals represent outgroups.
doi:10.1371/journal.pone.0024749.t004

No ancestral state for Aphidini (node A in Figure 3) was highly favored from among the three states: monoecy on trees, heteroecy, and monoecy on grasses. However, the two monoecious states were combined, monoecy had a higher probability than did heteroecy. Two nodes, B and C, in Rhopalosiphina showed near half proportions of heteroecy, 0.46 and 0.55, respectively, with regards to the origin of host alternation. *Melanaphis* aphids exhibit both heteroecy and monoecy on grasses. When the ancestral state of *Melanaphis japonica* was set to heteroecy, the proportion of the host alternation at node B increased to 0.51. However, the ancestral state of Aphidina (node D) was more likely monoecy on grasses (0.42) or monoecy on trees (0.36) than heteroecy (0.23). In general, ancestral host alternation was inferred to be less likely within Aphidina (nodes E, F, and H-L; 0.08–0.25), except for the clade of *Bursaphis* (node G; 0.55). In addition, more recent groups

(nodes H-L) within *Aphis* were highly inferred (0.76–0.84) to have originated from an ancestor that was monoecious holocyclic on herbaceous plants. Thus, the ancestral state of Rhopalosiphina seemed to be equivocal between heteroecy and monoecy, whereas that of Aphidina seemed to be monoecy.

Biogeographic origins

The origin of Aphidini was not clearly inferred to one region; both the European and Australasian regions received relatively high probabilities of 0.38 and 0.29, respectively (node A in Figure 4). Within Rhopalosiphina, the exact distributional origins also could not be predicted at nodes B and C, but the European region had the highest probabilities of 0.31 and 0.38, respectively, among all regions. The probability of an Australasian origin for Aphidina (node D) was 0.34, probably due to the basality of

Table 5. Divergence times or earliest fossil occurrences of host plants for Aphidinae and Pterocommatinae aphids.

Host-plant taxon	Epoch	Age (MYA) ^a	Method ^b	Related aphid taxon (node no.)	References
<i>Artemisia</i>	Middle to Late Miocene	10–23 ^c	Fossil	<i>Aphis kurosawai</i> (62)	[112]
Araliaceae	Middle Eocene	41–44	Dating (N)	P-C group (5)	[89]
Asteraceae	Early to Middle Eocene	51	Dating (N)	<i>Aphis terricola</i> (22)	[92]
Asteraceae	Early Eocene	42–48	Dating (P)	<i>Aphis terricola</i> (22)	[93]
<i>Centaurea</i> (Cardueae) ^d	Late Eocene	35–38	Dating (N)	<i>Aphis terricola</i> (22)	[92]
<i>Coprosma</i> (Anthospermeae) ^d	Middle Oligocene to Middle Eocene	31.8–47.6	Dating (B)	<i>Aphis coprosmae</i> (23)	[94]
<i>Epilobium</i> ^e and <i>Oenothera</i>	Oligocene to Miocene	12–35	Dating (P)	<i>Bursaphis</i> (31)	[99,100]
<i>Glycine</i>	Early to Middle Miocene	14.2–19.2	Dating (P)	<i>Aphis glycinis</i> (37)	[97]
<i>Glycine</i>	Early to Middle Miocene	8.0–11.0	Dating (P)	<i>Aphis glycinis</i> (37)	[98]
<i>Ligustrum</i> (Oleaceae) ^d	Middle Paleocene to Late Eocene	37–64	Dating (N)	<i>Aphis crinosa</i> (24)	[89]
<i>Miscanthus</i> (Paniceae) ^d	Early Miocene	20.6	Dating (M)	<i>Melanaphis japonica</i> (11)	[101]
<i>Phragmites</i>	Early Miocene	17.8–20.6	Dating (M)	<i>Hyalopterus pruni</i> (12)	[101]
<i>Prunus</i>	Middle Eocene	48 ^f	Fossil	<i>Hyalopterus pruni</i> (12)	[11]
<i>Prunus</i>	Middle Eocene	35	Dating (N)	<i>Hyalopterus pruni</i> (12)	[89]
<i>Rhamnus</i> (including <i>Frangulae</i>)	Late Oligocene	26.5–27.4	Dating (N)	<i>gossypii</i> group (35)	[95]
<i>Ribes</i> (Saxifragaceae) ^d	Late Cretaceous	89–96	Dating (P)	<i>Bursaphis</i> (31)	[96]
<i>Ribes</i> (Saxifragaceae) ^d	Late Cretaceous	81	Dating (N)	<i>Bursaphis</i> (31)	[89]
Rosaceae	Middle Eocene	35–56 ^c	Fossil	Aphidini+Macrosiphini (2, 3, 4)	[91]
Rosaceae	Late Cretaceous to Middle Eocene	44–76	Dating (N)	Aphidini+Macrosiphini (2, 3, 4)	[89]
Salicaceae	Middle Eocene	45 ^f	Fossil	P-C group (5)	[11]
<i>Salix</i> and <i>Populus</i>	Paleocene	60–65	Fossil	Aphidinae+P-C group (1)	[90]
Spiraeoideae (= Prunoideae)	Middle Miocene to Early Oligocene	29–44	Dating (N)	<i>Hyalopterus pruni</i> (12)	[89]

^a: geological time scale from Gradstein and Ogg [78].

^b: fossil means the earliest fossil record. Dating method in parenthesis: B, relaxed-clock in BEAST; M, relaxed-clock in MULTIDIVTIME; N, non-parametric rate smoothing in R8S; P, penalized likelihood in R8S.

^c: range of the epoch period.

^d: age inferred from the higher taxon.

^e: inferred by phylogenetic relationships with *Fuchsia* and *Oenothera*.

^f: absolute age based on the earliest fossil by Lopez-Vaamonde et al. [11].

doi:10.1371/journal.pone.0024749.t005

certain Southern Hemisphere species, although the European origin still constituted the largest proportion at 0.38. Thus, Aphidina probably diverged into European and Australasian lineages early in its evolution. Subsequent to that, large proportions of European ancestral origin were highly inferred for both nodes F and H, which radiated to the Nearctic region (node G) and subsequently to the Asian region (node I) at 28–33 MYA. Correspondingly, most Asian species originated from the MRCA of the *gossypii* group (nodes I and J), whereas the *craccivora*, *fabae*, and *spiraecola* groups more likely originated from the European ancestor (nodes K and L). Based on these results, morphological separation between the species-groups and morphological stasis within each species-group [27] may be caused by the regional isolation of the two conspicuous lineages that originated in the European and Asian regions.

Discussion

Phylogenetic relationships of Aphidini, Macrosiphini, and Pterocommatinae

The phylogeny presented in this study shows that the P-C group containing two genera, *Pterocomma* and *Cavariella*, is the most basal group of Aphidinae (Figure 1). Indeed, *Cavariella* should be

transferred into Pterocommatinae, because these groups share two common features: 1) primary host association with Salicaceae [2,14] and 2) morphological characteristics of fundatrices that are almost identical to their offspring [102,103]. Our phylogeny is consistent with phylogenies based on morphological characters, retaining the independent subfamily of Pterocommatinae [24,84,85]. The Pterocommatinae diverged early from the Aphidinae, and then Aphidinae diverged into Macrosiphini and Aphidini more recently.

In an earlier study based on a combination of two gene regions, tRNA/COII and EF1 α , von Dohlen et al. [21] suggested that the P-C group had a sister group relationship with Macrosiphini. Although the KH and AU tests in this study did not reject the two alternative topologies, i) ([P-C group+Aphidini]+Macrosiphini) and ii) ([P-C group+Macrosiphini]+Aphidini), the confidence values of both tests were approximately one-fifth of the best topology of (P-C group+[Aphidini+Macrosiphini]). Furthermore, the P-C group had relatively large genetic distances from Aphidini and Macrosiphini and also exhibited a long-branch from the root in both the ML and BI analyses. Our phylogeny is also consistent with the recent phylogeny by Ortiz-Rivas and Martinez Torres [36], in which two nuclear genes, long-wave length opsin and ATP6 (1,360 bp), were used together with tRNA/COII and

EF1 α . Our study based on an alternative combination of mitochondrial and nuclear sequences also supports the basality of the P-C group.

The monophyly of both Aphidina and Rhopalosiphina was well supported and corresponded to previous phylogenies of Aphidini [25,26]. Although five DNA regions (4,500 bp) were used for the analyses in this study, the monophyly of *Rhopalosiphum*, *Melanaphis*, and *Schizaphis* was not resolved. Moreover, two *Melanaphis* species did not form a clade as they adapted to two unrelated plant genera, *Micanthus* and *Luzula* [14]. In *Aphis*, however, each of four species-groups and the subgenus *Buraphis* clearly formed monophyly, even though the Aphidina species were genetically closer to one another than the Rhopalosiphina species. The inconsistencies between the taxonomic and phylogenetic relationships are likely caused by faulty diagnoses for the genera of Rhopalosiphina [84,85]. Thus, the generic division and classification within Rhopalosiphina need to be revised.

Evolution of host plant association and host alternation

Although molecular dating remains controversial due to different molecular rates across lineages [104], this technique is widely used for phylogenetic reconstruction and determining evolutionary patterns [10,11]. The estimation of divergence times suggests that aphid taxa used in this study likely diversified during the radiation period of their host plants (Table 4, 5). The diversification periods of aphid taxa and their hosts were overlapping, even though the divergence time estimates for hosts differed depending on which dating methods and fossil information were used (Figure 2; Table 5). However, the time estimates could not give an explanation for topological coincidences of co-diversification [22] due to the promiscuous host association in Aphidini [21,25,26].

The most striking result from this study was that extant heteroecious species could not use their secondary hosts before the Oligocene, because their secondary hosts emerged between the Oligocene and Miocene (Table 5). In other words, there were large temporal differences between the occurrences of primary and secondary hosts. As von Dohlen et al. [21] discussed, secondary hosts such as grasses and dicotyledonous herbs were not the major elements of temperate plant communities in the North Hemisphere, at least until the Miocene. The host association of *Melanaphis* can be viewed as crucial evidence since the earliest origins of C4 grasses, including *Miscanthus*, likely occurred about 32 MYA during the Oligocene [101]. Moreover, most heteroecious species in *Rhopalosiphum* and *Schizaphis* have adapted to many C4 grasses as a secondary host, such as *Echinochloa*, *Panicum*, *Pennisetum*, *Setaria*, *Sorghum*, and *Zea*, and they also utilize relatively young C3 grasses such as *Phragmites* and *Oryza* [14,101]. The divergence times of heteroecious aphid genera were more congruent with the diversification of the primary hosts. It is tempting to conclude that the origin of species-level diversification coincided with the occurrences of the secondary hosts. However, more studies across diverse genera are needed to generalize the association of the species-level diversification in heteroecious aphids and their secondary hosts.

Our study also supports the multiple origins of host alternation [4,16] within Aphidini. Von Dohlen [21] suggested that host alternation originated independently from Pterocommatinae, Macrosiphini, and Aphidini. However, our results differ from this basic premise in that Aphidini might have originated from monoecious ancestors. In the basal positions of Aphidina, three *Protaphis*-like species, *Aphis terricola*, *A. coprosmae*, and *A. crinosa*, likely diverged before the Oligocene and are monoecious with holocycle [14]. Instead, host alternation evolved further down the phylogeny

independently in Rhopalosiphina and Aphidina. The likelihood of host-alternating origins in Aphidini (Figure 3; nodes A) is very low, whereas the group alternating between *Prunus* and Poaceae (node C) and the group alternating between *Ribes* and Onagraceae (node G) had a likelihood over 0.5 for host alternation, which diverged after the Middle Oligocene. In addition, *Cavariella* might have acquired host alternation earlier as seen in its time of diversification, which is consistent with that of Salicaceae and Araliaceae or closely-related Apiaceae occurring in the Middle to Late Eocene. Thus, there might be at least four independent origins of host alternation in Pterocommatinae+Aphidinae in our study.

The hypothesis of multiple origins of host alternation could conflict with the idea of the partitioned migration of females and males in Aphidinae, i.e., a single origin of separate migration of sexual winged males and females (i.e., gynoparae) bearing wingless egg-laying females (i.e., oviparae) (see von Dohlen et al. [21]). However, von Dohlen et al. [21] suggested that partitioning of winged male versus wingless oviparous female embryos into different viviparous females could be a plesiomorphic trait for Aphidinae. Although the gynopara is specialized to return to its primary host using its sensory capabilities [105], it is still uncertain whether this morph is the evolutionary result of host alternation. Except for two generations required to produce sexual females, mating between winged males and wingless oviparous females produced by winged viviparous migrants (i.e., sexuparae) occurs in other related monoecious taxa, including Calaphidinae, Chitophorinae, Drepanosiphinae, and Lachninae [2,36]. In this light, the separate migration was likely acquired upon divergence from these monoecious subfamilies, as *Cavariella* also has gynoparae [21]. However, the other host-alternating aphid groups (Anoeciinae, Eriosomatinae, and Hormaphidinae), which are apparently phylogenetically distant from Aphidinae [36], still have sexuparae that produce both male and female sexuals in the primary host [3]. Therefore, one wonders why these aphids have not evolved specialized gynoparae like those in Aphidinae, even though most species in these groups alternate primary and secondary hosts in a one-year life cycle [4]. It might be concluded either that the separate migration is either plesiomorphic in Aphidinae as a whole or is unrelated to host alternation, in which case the life-history trait of host alternation most likely arose several times within Aphidinae.

Possible selection pressures for the evolution of host alternation in the Oligocene included climate change [106,107] in conjunction with the nutritional superiority of herbaceous hosts [102,108]. The origin of host alternation likely occurred during the rise of secondary herbaceous plants after the Oligocene climate change [101]. During the Oligocene, temperature and CO₂ levels decreased, shifting the global climate toward more arid conditions [106]. Herbaceous plants such as the C4 grasses became dominant [101]. Aphids alternating their primary and secondary hosts might have obtained better nutritional sources [102,109]. Alternatively, evolution of host alternation could be explained by fundatrix constraint, enemy escape, bet hedging, or induced responses hypotheses [110]. Unlike the climate change hypothesis, these hypotheses explain the maintenance of host alternation in aphids, rather than the origin or timing of the host alternation trait.

Biogeography of the Aphidini

The biogeographic origin of the three main lineages of aphids (nodes A, B, and D) were equivocal, although a European origin received the highest probabilities in these three nodes (Figure 4). Within Rhopalosiphina, geographic inconsistencies made the species within each clade unclear. However, within Aphidina,

four lineages (F, H, K, L) likely originated in Europe, whereas the *gossypii* group (node I) was most likely of an Asian origin. This suggests that Asian endemic species (e.g., *Aphis clerodendri*, *A. egomae*, and *A. sumire*) most likely originated from a common ancestor of node I (which includes the *gossypii* group), whereas European endemic species (including the *fabae+craccivora+spiraeicola* groups) likely originated from a common ancestor of node K. In addition, *Bursaphis* originated in the Nearctic (node G), and the four Southern Hemisphere species originated from Australasia after diverging from the European lineage. Interestingly, based on this result, the classical morphological groups of *Aphis* (i.e., subgenera and *Aphis* species-groups) were possibly separated by geographic isolation within Aphidina [14,26,27,30,84,85,111].

Although the nodes received low likelihood scores, the European and Australasian regions were more likely the biogeographic origins for Aphidini than were the Asian or Nearctic regions. It seems that *Aphis terricola* diverged earliest among all aphidine aphids within Aphidini, and all five Southern Hemisphere species diverged relatively early within Aphidina. This inference hinges largely on the geographic origin of the basal taxa. No extant species diverged earlier than the subgenus *Protaphis* (*A. (P.) terricola*) within Aphidini, based on morphological and molecular systematics [30,84,85]. Furthermore, the two other *Protaphis*-like species, *A. coprosmae* and *A. crinosa*, subsequently diverged after the divergence of *A. terricola*, and the genus *Melanaphis* resembles many aphids in *Protaphis*, which is the basal group of Rhopalosiphina [26,84,85,88]. It is rather interesting that the three species, *A. terricola*, *A. coprosmae*, and *A. crinosa*, endemic to separate regions, appear in basal positions on the phylogeny in spite of the geographic gaps between them [14,88]. To confirm the biogeographic origin of aphidine aphids, more research on phylogenies including more *Protaphis* and *Protaphis*-like species should be performed.

References

- Dixon AFG (1987) The way of life of aphids: host specificity, speciation and distribution. In: Minks AK, Harrewijn P, eds. Aphids their biology, natural enemies and control, Vol A. Amsterdam: Elsevier. pp 197–207.
- Blackman RL, Eastop VF (1994) Aphids on the World's Trees: An Identification and Information Guide. Wallingford: CAB International. 987 p.
- Blackman RL, Eastop VF (2000) Aphids on the World's Crops: An Identification and Information Guide. Chichester: John Wiley & Sons Ltd. 466 p.
- Moran NA (1988) The evolution of host-plant alternation in aphids: Evidence for specialization as a dead end. *The American Naturalist* 132: 681–706.
- Moran NA (1992) The evolution of aphid life cycles. *Annual Review of Entomology* 37: 321–348.
- Mitter C, Farrell B, Futuyma D (1991) Phylogenetic studies of insect-plant interactions: Insights into the genesis of diversity. *Trends in Ecology and Evolution* 6: 290–293.
- Farrell BD, Mitter C (1998) The timing of insect/plant diversification: Might *Tetraopes* (Coleoptera: Cerambycidae) and *Asclepias* (Asclepiadaceae) have co-evolved? *Biological Journal of the Linnean Society* 63: 553–577.
- Weiblen GD (2001) Phylogenetic relationships of fig wasps pollinating functionally dioecious *Ficus* based on mitochondrial DNA sequences and morphology. *Systematic Biology* 50: 243–267.
- Futuyma DJ, Agrawal AA (2009) Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences of the United States of America* 106: 18054–18061.
- Lopez-Vaamonde C, Wikström N, Kjer KM, Weiblen GD, Rasplus JY, et al. (2009) Molecular dating and biogeography of fig-pollinating wasps. *Molecular Phylogenetics and Evolution* 52: 715–726.
- Lopez-Vaamonde C, Wikström N, Labandeira C, Godfray HCJ, Goodman SJ, et al. (2006) Fossil-calibrated molecular phylogenies reveal that leaf-mining moths radiated millions of years after their host plants. *Journal of Evolutionary Biology* 19: 1314–1326.
- Lopez-Vaamonde C, Rasplus JY, Weiblen GD, Cook JM (2001) Molecular phylogenies of fig wasps: Partial cocladogenesis of pollinators and parasites. *Molecular Phylogenetics and Evolution* 21: 55–71.
- Percy DM, Page RDM, Cronk QCB (2004) Plant-insect interactions: Double-dating associated insect and plant lineages reveals asynchronous radiations. *Systematic Biology* 53: 120–127.
- Blackman RL, Eastop VF (2006) Aphids on the World's Herbaceous Plants and Shrubs, Vol. 2, The Aphids. Chichester: John Wiley & Sons Ltd. pp 1025–1439.
- Heie OE (1987) Paleontology and phylogeny. In: Minks AK, Harrewijn P, eds. Aphids Their Biology, Natural Enemies and Control, vol A. Amsterdam: Elsevier. pp 367–391.
- von Dohlen CD, Moran NA (2000) Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biological Journal of the Linnean Society* 71: 689–717.
- Favret C (2011) Aphid Species File. Version 1.0/4.0. Available at: <http://Aphid.SpeciesFile.org>. Accessed 2011 Mar 11.
- Wojciechowski W (1992) Studies on the Systematic System of Aphids (Homoptera, Aphidinea). Katowice: Uniwersytet Slaski.
- Heie OE (1994) Aphid ecology in the past and a new view on the evolution of Macrosiphini. In: Leather SR, ed. Individuals, Populations and Patterns in Ecology. Andover, Hampshire, UK: Intercept Ltd. pp 409–418.
- Moran NA (1990) Aphid life cycles: Two evolutionary steps. *The American Naturalist* 136: 135–138.
- von Dohlen CD, Rowe CA, Heie OE (2006) A test of morphological hypotheses for tribal and subtribal relationships of Aphidinae (Insecta: Hemiptera: Aphididae) using DNA sequences. *Molecular Phylogenetics and Evolution* 38: 316–329.
- Havill NP, Footitt RG, von Dohlen CD (2007) Evolution of host specialization in the Adelgidae (Insecta: Hemiptera) inferred from molecular phylogenetics. *Molecular Phylogenetics and Evolution* 44: 357–370.
- Footitt RG, Maw HEL, Von Dohlen CD, Hebert PDN (2008) Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. *Molecular Ecology Resources* 8: 1189–1201.
- Remaudiere G, Remaudiere M (1997) Catalogue des Aphididae du Monde. Homoptera Aphidoidea; Catalogue of the world's Aphididae. Paris: INRA. 473 p.
- von Dohlen CD, Teulon DAJ (2003) Phylogeny and historical biogeography of New Zealand indigenous aphidini aphids (Hemiptera, Aphididae): An hypothesis. *Annals of the Entomological Society of America* 96: 107–116.
- Kim H, Lee S (2008) A molecular phylogeny of the tribe Aphidini (Insecta: Hemiptera: Aphididae) based on the mitochondrial tRNA/COII, 12S/16S and the nuclear EF1 α genes. *Systematic Entomology* 33: 711–721.

Supporting Information

Figure S1 Calibration points used as fixation or constraint of nodes for estimation of divergence times.

(EPS)

Table S1 Aphid species used in this study with GenBank accession numbers, voucher numbers, and reference. Classification following Remaudière and Remaudière [24].

(DOC)

Table S2 Primers used for DNA amplification and sequencing.

(DOC)

Table S3 Life cycles, host types, and host plants of the Aphidini aphids.

(DOC)

Table S4 Biogeographic origin of the Aphidini aphids.

(DOC)

Table S5 Estimated divergence times of all nodes. The node numbers are those of Figures 1 and 2.

(DOC)

Acknowledgments

We would like to thank Dr. Roger L. Blackman of the Natural History Museum, London, U.K., for comments on the most primitive taxa in Aphidini, Dr. Carol D. von Dohlen of the Utah State University, Utah, U.S.A., for criticisms on an earlier draft of the manuscript, and Dr. Ole E. Heie and Dr. Sebastiano Barbagallo for comments on the ancestral host candidates of the *gossypii* group at the 8th ISA meeting.

Author Contributions

Conceived and designed the experiments: HK SL YJ. Performed the experiments: HK. Analyzed the data: HK. Contributed reagents/materials/analysis tools: HK SL. Wrote the paper: HK YJ.

27. Kim H, Lee W, Lee S (2010) Morphometric relationship, phylogenetic correlation, and character evolution in the species-rich genus *Aphis* (Hemiptera: Aphididae). *PLoS One* 5: e11608.
28. Kim H, Hoelmer KA, Lee W, Kwon YD, Lee S (2010) Molecular and morphological identification of the soybean aphid and other *Aphis* species on the primary host *Rhannus daurica* in Asia. *Annals of the Entomological Society of America* 103: 532–543.
29. Turcinaviciene J, Rakauskas R, Pedersen BV (2006) Phylogenetic relationships in the “*gossulariae*” species group of the genus *Aphis* (Hemiptera: Sternorrhyncha: Aphididae): Molecular evidence. *European Journal of Entomology* 103: 597–604.
30. Coeur d’acier A, Jousset E, Martin JF, Rasplus JY (2007) Phylogeny of the genus *Aphis* Linnaeus, 1758 (Homoptera: Aphididae) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 42: 598–611.
31. Thao ML, Baumann L, Baumann P (2004) Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evolutionary Biology* 4: 25.
32. Carletto J, Blin A, Vanlerberghe-Masutti F (2009) DNA-based discrimination between the sibling species *Aphis gossypii* Glover and *Aphis frangulae* Kaltensch. *Systematic Entomology* 34: 307–314.
33. Yang ZX, Chen XM, Havill NP, Feng Y, Chen H (2010) Phylogeny of *Rhus* gall aphids (Hemiptera: Pemphigidae) based on combined molecular analysis of nuclear EF1 alpha and mitochondrial COII genes. *Entomological Science* 13: 351–357.
34. Qiao G, Wang J, Zhang G (2008) *Toxoptera* Koch (Hemiptera: Aphididae), a generic account, description of a new species from China, and keys to species. *Zootaxa* 1746: 1–14.
35. Martinez-Torres D, Buades C, Latorre A, Moya A (2001) Molecular systematics of aphids and their primary endosymbionts. *Molecular Phylogenetics and Evolution* 20: 437–449.
36. Ortiz-Rivas B, Martinez-Torres D (2010) Combination of molecular data support the existence of three main lineages in the phylogeny of aphids (Hemiptera: Aphididae) and the basal position of the subfamily Lachninae. *Molecular Phylogenetics and Evolution* 55: 305–317.
37. Moran NA (1989) A 48-million-year-old aphid-host plant association and complex life cycle: biogeographic evidence. *Science* 245: 173–175.
38. Moran NA, Munson MA, Baumann P, Ishikawa H (1993) A Molecular Clock in Endosymbiotic Bacteria is Calibrated Using the Insect Hosts. *Proceedings of the Royal Society of London Series B-Biological Sciences* 253: 167–171.
39. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3: 294–299.
40. Stern DL (1994) A phylogenetic analysis of soldier evolution in the aphid family Hormaphididae. *Proceedings of the Royal Society of London Series B-Biological Sciences* 256: 203–209.
41. Normark BB (1996) Phylogeny and evolution of parthenogenetic weevils of the *Aramigus tessellatus* species complex (Coleoptera: Curculionidae: Naupactini): Evidence from mitochondrial DNA sequences. *Evolution* 50: 734–745.
42. Harry M, Solignac M, Lachaise D (1998) Molecular evidence for parallel evolution of adaptive syndromes in fig-breeding *Lissocephala* (Drosophilidae). *Molecular Phylogenetics and Evolution* 9: 542–551.
43. Jermin L, Crozier RH (1994) The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: sequence divergence in Hymenoptera may be associated with nucleotide content. *Journal of Molecular Evolution* 38: 282–294.
44. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, et al. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. *Annals of the Entomological Society of America* 87: 651–701.
45. Simon C, Franke A, Martin A (1991) The polymerase chain reaction: DNA extraction and application. In: Hewitt GM, Johnston AWB, Young JPW, eds. *Molecular Techniques in Taxonomy*. Berlin: Springer. 410 p.
46. Palumbi SR (1996) Nucleic acids II: The polymerase chain reaction. In: Hillis DM, ed. *Molecular Systematics*. Sunderland: Sinauer Press. pp 205–247.
47. von Dohlen CD, Kurosu U, Aoki S (2002) Phylogenetics and evolution of the eastern Asian-eastern North American disjunct aphid tribe, Hormaphidini (Hemiptera: Aphididae). *Molecular Phylogenetics and Evolution* 23: 257–267.
48. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876–4882.
49. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 9: 299–306.
50. Castresana J (2002) GBLOCKS: Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Version 0.91b. Copyrighted by J. Castresana, EMBL.
51. Swofford DL (1998) PAUP*. *Phylogenetic Analysis Using Parsimony (* and Other Methods)*, Version 4.0b10. Sunderland, MA: Sinauer Associates.
52. Farris JS, Källersjö M, Kluge AG, Bult C (1994) Testing significance of incongruence. *Cladistics* 10: 315–319.
53. Nylander J (2004) MrModeltest 2.0. Program distributed by the author. Uppsala university, Evolutionary biology centre.
54. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
55. Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53: 793–808.
56. Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253–1256.
57. Stamatakis A (2006) RAXML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
58. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
59. Rambaut A, Drummond AJ (2009) Tracer v1.5. Available at: <http://beast.bio.ed.ac.uk/tracer>. Accessed 2011 Mar 11.
60. Dunn KA, McEachran JD, Honeycutt RL (2003) Molecular phylogenetics of myliobatiform fishes (Chondrichthyes: Myliobatiformes), with comments on the effects of missing data on parsimony and likelihood. *Molecular Phylogenetics and Evolution* 27: 259–270.
61. Wiens JJ (2003) Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology* 52: 528–538.
62. Wiens JJ (2003) Incomplete taxa, incomplete characters, and phylogenetic accuracy: Is there a missing data problem? *Journal of Vertebrate Paleontology* 23: 297–310.
63. Philippe H, Snell EA, Baptiste E, Lopez P, Holland PWH, et al. (2004) Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Molecular Biology and Evolution* 21: 1740–1752.
64. Fulton TL, Strobeck C (2006) Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. *Molecular Phylogenetics and Evolution* 41: 165–181.
65. Boučeknak-Khelladi Y, Salamin N, Savolainen V, Forest F, van der Bank M, et al. (2008) Large multi-gene phylogenetic trees of the grasses (Poaceae): Progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47: 488–505.
66. Wiens JJ (2005) Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Systematic Biology* 54: 731–742.
67. Wiens JJ (2006) Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* 39: 34–42.
68. Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution* 29: 170–179.
69. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51: 492–508.
70. Maddison WP, Maddison DR (2007) Mesquite: A modular system for evolutionary analysis. Version 2.6. Available at: <http://mesquiteproject.org>. Accessed 2011 Mar 11.
71. Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13: 555–556.
72. Shimodaira H, Hasegawa M (2001) CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.
73. Heie OE (1994) The Aphidoidea (Hemiptera) of Fennoscandia and Denmark. V. Family Aphididae: Part 2 of tribe Macrosiphini of subfamily Aphidinae, and family Lachninae. Leiden, Netherlands: E.J. Brill/Scandinavian Science Press Ltd. 242 p.
74. Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. *Molecular Biology and Evolution* 16: 1586–1598.
75. Kishino H, Thorne JL, Bruno WJ (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution* 18: 352–361.
76. Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* 51: 689–702.
77. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.
78. Gradstein FM, Ogg JG (2004) Geologic Time Scale 2004 - Why, how, and where next! *Lethaia* 37: 175–181.
79. Rutschmann F (2005) Bayesian molecular dating using PAML/multidivtime. A step-by-step manual. University of Zurich, Switzerland. Available at <http://www.plant.ch>. Accessed 2011 Mar 11.
80. Felsenstein J (1984) DNAML in PHYLIP 2.6. University of Washington, Seattle.: Available at: <http://evolution.genetics.washington.edu>. Accessed 2011 Mar 11.
81. Rambaut A (2009) FigTree v1.3.1. Available at: <http://tree.bio.ed.ac.uk/software/figtree>. Accessed 2011 Mar 11.
82. Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53: 673–684.
83. Pagel M, Meade A (2007) BayesTraits, version 1.0 - Draft Manual. Available at: <http://www.evolution.rdg.ac.uk>. Accessed 2011 Mar 11.
84. Stroyan HLG (1984) Aphids-Pterocommatinae and Aphidinae (Aphidini) Homoptera, Aphididae. *Handbk. Ident. Br. Insects* 2, pt. 6. London: Dramrite Printers Ltd.
85. Heie OE (1986) The Aphidoidea (Hemiptera) of Fennoscandia and Denmark. III. Family Aphididae: subfamily Pterocommatinae & tribe Aphidini of subfamily Aphidinae. Klampenborg, Denmark: Scandinavian Science Press Ltd.

86. Teulon DAJ, Stufkens MAW (1998) Current status of New Zealand indigenous aphids. Wellington, New Zealand: Department of Conservation. pp 1–23.
87. Lee S, Holman J, Havelka J (2002) Illustrated Catalogue of Aphididae in the Korean Peninsula. Part I, Subfamily Aphidinae (Hemiptera: Sternorrhyncha). Daejeon, Rep. of Korea: Korea Research Institute of Bioscience and Biotechnology. 329 p.
88. Lee S, Kim H (2006) Economic Insects of Korea 28 (Insecta Koreana Suppl. 35), Aphididae: Aphidini (Hemiptera: Sternorrhyncha). Suwon, Rep. of Korea: National Institute of Agricultural Science and Technology.
89. Wikström N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: Calibrating the family tree. Proceedings of the Royal Society of London Series B-Biological Sciences 268: 2211–2220.
90. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596–1604.
91. Taylor TN, Taylor EL, Krings M (2009) Paleobotany, The biology and evolution of fossil plants. London: Academic Press, Ltd.
92. Kim KJ, Choi KS, Jansen RK (2005) Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). Molecular Biology and Evolution 22: 1783–1792.
93. Bremer K, Friis EM, Bremer B (2004) Molecular phylogenetic dating of asterid flowering plants shows early Cretaceous diversification. Systematic Biology 53: 496–505.
94. Bremer B, Eriksson T (2009) Time tree of Rubiaceae: Phylogeny and dating the family, subfamilies, and tribes. International Journal of Plant Sciences 170: 766–793.
95. Richardson JE, Chatrou LW, Mols JB, Erkens RHJ, Pirie MD (2004) Historical biogeography of two cosmopolitan families of flowering plants: Amnaceae and Rhamnaceae. Philosophical Transactions of the Royal Society B-Biological Sciences 359: 1495–1508.
96. Jian SG, Soltis PS, Gitzendanner MA, Moore MJ, Li R, et al. (2008) Resolving an ancient, rapid radiation in Saxifragales. Systematic Biology 57: 38–57.
97. Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Systematic Biology 54: 575–594.
98. Stefanovic S, Pfeil BE, Palmer JD, Doyle JJ (2009) Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. Systematic Botany 34: 115–128.
99. Ford VS, Gottlieb LD (2007) Tribal relationships within Onagraceae inferred from *PgiC* sequences. Systematic Botany 32: 348–356.
100. Rutschmann F, Eriksson T, Abu Salim K, Conti E (2007) Assessing calibration uncertainty in molecular dating: The assignment of fossils to alternative calibration points. Systematic Biology 56: 591–608.
101. Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA (2008) The age of the grasses and clusters of origins of C-4 photosynthesis. Global Change Biology 14: 2963–2977.
102. Mackenzie A, Dixon AFG (1990) Host alternation in aphids: Constraint versus optimization. The American Naturalist 136: 132–134.
103. Heie OE (1992) The Aphidoidea (Hemiptera) of Fennoscandia and Denmark. IV. Family Aphididae: Part 1 of tribe Macrosiphini of subfamily Aphidinae. Leiden, Netherlands: E.J. Brill/Scandinavian Science Press Ltd.
104. Parham JF, Irmis RB (2008) Caveats on the use of fossil calibrations for molecular dating: A comment on near et al. The American Naturalist 171: 132–136.
105. Powell G, Tosh CR, Hardie J (2006) Host plant selection by aphids: Behavioral, evolutionary, and applied perspectives. Annual Review of Entomology 51: 309–330.
106. Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292: 686–693.
107. Yamamoto S, Sota T (2007) Phylogeny of the Geometridae and the evolution of winter moths inferred from a simultaneous analysis of mitochondrial and nuclear genes. Molecular Phylogenetics and Evolution 44: 711–723.
108. Mackenzie A (1996) A trade-off for host plant utilization in the black bean aphid, *Aphis fabae*. Evolution 50: 155–162.
109. Kundu R, Dixon AFG (1995) Evolution of complex life cycles in aphids. Journal of Animal Ecology 64: 245–255.
110. Havill NP, Footitt RG (2007) Biology and evolution of Adelgidae. Annual Review of Entomology 52: 325–349.
111. Holman J (1987) Notes on *Aphis* species from the Soviet Far East, with descriptions of eight new species (Homoptera, Aphididae). Acta Entomologica Bohemoslovaca 84: 353–387.
112. Wang HC, Moore MJ, Soltis PS, Bell CD, Brockington SF, et al. (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. Proceedings of the National Academy of Sciences of the United States of America 106: 3853–3858.