



HAL
open science

Discordance between mitochondrial, nuclear, and symbiont genomes in aphid phylogenetics: who is telling the truth?

Emmanuelle Jouselin, Armelle Coeur d'Acier, Anne-laure Clamens, Maxime Galan, Corinne Cruaud, Valerie Barbe, Alejandro Manzano-Marín

► To cite this version:

Emmanuelle Jouselin, Armelle Coeur d'Acier, Anne-laure Clamens, Maxime Galan, Corinne Cruaud, et al.. Discordance between mitochondrial, nuclear, and symbiont genomes in aphid phylogenetics: who is telling the truth?. *Zoological Journal of the Linnean Society*, 2024, 201 (4), pp.zlae098. 10.1093/zoolinnean/zlae098 . hal-04692066

HAL Id: hal-04692066

<https://hal.inrae.fr/hal-04692066v1>

Submitted on 9 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.



L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Original Article

Discordance between mitochondrial, nuclear, and symbiont genomes in aphid phylogenetics: who is telling the truth?

Emmanuelle Jouselin¹, Armelle Coeur d'acier¹, Anne-Laure Clamens¹, Maxime Galan¹,
Corinne Cruaud² , Valérie Barbe², Alejandro Manzano-Marín³ 

¹UMR 1062 Centre de Biologie pour la Gestion des Populations, INRAE, CIRAD, Institut Agro Montpellier, IRD, Univ. Montpellier. 755 avenue du Campus Agropolis, 34988 Montferrier-sur-Lez, France

²Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay. 2 Rue Gaston Crémieux, 91057 Evry, France

³Centre for Microbiology and Environmental Systems Science, University of Vienna. Djerassiplatz 1, 1030 Vienna, Austria

Corresponding authors: Centre for Microbiology and Environmental Systems Science, University of Vienna. Djerassiplatz 1, 1030 Vienna, Austria.

E-mail: alejandro.manzano.marin@univie.ac.at; UMR 1062 Centre de Biologie pour la Gestion des Populations, INRAE, CIRAD, Institut Agro Montpellier, IRD, Univ. Montpellier. 755 avenue du Campus Agropolis, 34988 Montferrier-sur-Lez, France. E-mail: emmanuelle.jouselin@inrae.fr

ABSTRACT

Aphids (Aphididae) are intensively studied due to their significance as pests and their captivating biological traits. Despite this interest, the evolutionary history of this insect family is poorly understood. Recent phylogenomic analyses have produced conflicting topologies, complicating our understanding of aphid trait evolution. In this work, we aimed to unravel the backbone phylogeny of aphids. We sequenced partial and whole mitochondrial genomes from 87 species. We additionally sequenced 42 nuclear loci across 95 aphid species and sourced 146 genes from 12 new and 61 published genomes from aphid obligate endosymbiont, *Buchnera aphidicola*. We obtained data from these three sources for a subset of aphid species, facilitating a comparative analysis of their signal. Our analyses confirm the monophyly of most subfamilies, validating current taxonomic classifications. However, relationships between subfamilies remain contentious in both mitochondrial and nuclear phylogenies. The topologies obtained with *Buchnera* are fully resolved but disagree with host phylogenies at deep evolutionary scales and conflict with views on the evolution of aphid morphology. We discuss alternative hypotheses for these discrepancies. Finally, the paucity of phylogenetic information at deep timescales may stem from an initial rapid radiation. Though challenging to establish, this scenario may inherently hinder resolution in aphid phylogenetics.

Keywords: Aphididae; endosymbiont; mitogenome; phylogenetic conflict; radiation; systematics

INTRODUCTION

Aphids (Hemiptera: Aphididae) constitute a lineage of over 5000 species of sap-feeding insects that include several significant agricultural, horticultural, and forestry pests (Blackman and Eastop 2000). This pest status has prompted numerous research programmes aimed at a better understanding of their biology for the development of appropriate control methods. Beyond their notoriety as pest insects, aphids also exhibit variable biological characteristics, making them a prominent study system in evolutionary biology. For instance, their patterns of associations with host plants range from strict specialization towards a single host-plant species to extreme polyphagy (with some aphid species capable of feeding on up to 100 plant families). Additionally, 10% of all aphid species undergo seasonal alternation between

two sets of host plants (Blackman and Eastop 2000, 2006), requiring the ability to locate and feed on plants belonging to different botanical families (Mackenzie and Dixon 1990, Powell and Hardie 2001, Jouselin *et al.* 2010). These features make aphids interesting systems for investigating how phytophagous insects adapt to their host plants and whether these adaptive processes drive speciation events (Shaposhnikov 1961, 1987, Heie 2004, Jouselin and Elias 2019). Aphids have also been studied intensively for their reproductive strategies undergoing cyclical parthenogenesis—alternating between sexual and asexual reproduction throughout their one or two year-long life cycles. Both the ecological significance and genetic determinism of this biological character have attracted the interest of evolutionary biologists (Delmotte *et al.* 2001, Jaquiere *et al.* 2014).

Received 14 April 2024; revised 14 June 2024; accepted 10 July 2024

© 2024 The Linnean Society of London.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Aphids also stand out as a study system in bacterial endosymbiosis. Since Buchner's pioneering work on endosymbiosis in invertebrates (Buchner 1965b), the study of their mutualistic association with their obligate nutritional symbiont, *Buchnera aphidicola*, has shed light on various aspects of the evolution of bacterial integration into eukaryotes. These aspects range from insect/symbiont co-speciation (Munson *et al.* 1991, Clark *et al.* 2000, Jousselin *et al.* 2009), metabolic cooperation (Wilson *et al.* 2010, Feng *et al.* 2019, Smith and Moran 2020) to degenerative evolution in endosymbiont genomes (Moran and Mira 2001, Gil *et al.* 2002, Toft and Fares 2009, Sabater-Muñoz *et al.* 2017). Many of the evolutionary biology studies cited above have been conducted on a few focal aphid species or groups. However, a thorough understanding of aphid biological evolution requires knowledge of how aphid traits are distributed throughout the diversification of this lineage. Generating a highly accurate and robust aphid phylogenetic tree is a crucial step to that effect.

In addition to the reconstruction of the history of aphid biological traits, aphid phylogenetic history could help in testing several macroevolutionary scenarios that have long been debated in the literature. The distribution patterns of aphid species through time and space is somewhat intriguing. The over 5000 currently described aphid species (Favret 2024) are organized into 23 subfamilies encompassing 510 genera. The first fossils identified as belonging to Aphididae date back to 160 million years ago (Mya) and the bulk of aphid fossil diversity is found in the Eocene (i.e. 56–33 Mya) (Heie 2004) and <https://paleobiodb.org/> (last accessed 14/12/2023). The 23 aphid subfamilies differ greatly in terms of number of species, with the most diverse subfamily, Aphidinae, gathering nearly half the diversity (with 2483 described species), while some species-poor subfamilies, such as Phloeomyzinae for instance, currently count no more than a single species. Differences in species' diversity could be explained by the different ages of the subfamilies, with depauperate subfamilies being more recent and, therefore, having had little time to diversify and accumulate species. Alternatively, these variations in aphid species' diversity could be caused by bursts of radiations and/or extinctions differentially affecting lineages of aphids. The latter scenario is clearly favoured by taxonomists and palaeontologists (Shaposhnikov 1981, Heie 1987, 2004) and early phylogenetic studies (von Dohlen and Moran 2000, Ortiz-Rivas *et al.* 2004). Another outstanding feature of aphids is their atypical geographic distribution pattern. They are highly diversified in the temperate region of the northern hemisphere (Heie 1994) and, therefore, show an inverse species' diversity distribution to other insect groups (that are generally more diverse in the tropical region; Brown 2014). Several hypotheses have been put forward to explain this pattern. These involve ecological factors, such as a hyper-specialization on a few host plants that might be maladapted to tropical environments (Dixon *et al.* 1987) or constraints linked to aphid historical biogeography (Shaposhnikov 1987). A thorough investigation of the scenarios underlying aphid species' distribution through time and space necessitates a robust phylogenetic reconstruction. Successive bursts of speciation and extinction could be traced back, and historical biogeography scenarios could be untangled.

Despite the wealth of scientific questions that necessitate a robust phylogenetic reconstruction, the aphid phylogenetic

tree remains unresolved. The debate over the evolutionary relationships among aphid species groups actually predates the advent of molecular phylogenetics. This is well illustrated by the numerous taxonomic classifications that have been proposed over the years. The number and delimitation of aphid groups and their ranks (family, subfamily, tribe) varied significantly among different authors. As Heie accurately pointed out, there were 'as many classifications as taxonomists' (Heie 1980). Ilharco and Harten (1987) in their review of aphid taxonomic classifications listed four to 21 subfamilies grouped in one to 10 families, depending on the authors. Wojciechowski (1992), in a more exhaustive review and an attempt at reconstructing a morphology-based phylogeny, cites and comments on 17 classifications in which delimitations of families varied. Wiczeorek and Wojciechowski (2007), focusing solely on the group of taxa known as Drepanosiphine aphids (mainly Drepanosiphidae *sensu* Heie 1980), noted that this group included three to 14 subfamilies, depending on classifications. These taxonomic controversies were succinctly outlined in the introduction by Remaudière and Stroyan (1984), prompting the formulation of a classification featuring 20 subfamilies within a family named Aphididae. This classification was conceived with a practical perspective, wherein each subfamily grouped species (tribes, genera) perceived as homogenous by all taxonomists. Homogeneity, in this context, was defined by the exhibition of morphological characters distinctive enough to unequivocally differentiate and categorize aphids in a group (whether the groups was seen as a subfamily or a tribe by previous studies). This classification, therefore, ranked some well-recognized taxa at a similar subfamily level. It was, therefore, less structured than previous classifications and, as underlined by Ilharco and Harten (1987), it made very few assumptions about the evolutionary trajectory of aphid characters. Attempts at higher level classification, grounded in intuitive assessments of aphid general morphology rather than formal cladistic analyses (Quednau 2010), have then been sporadic. Remaudière's (1984) classification gained popularity through the Aphid World Catalog (Remaudière and Remaudière 1997). Molecular studies have then often confirmed the monophyly of subfamilies (e.g. Chen *et al.* 2014, 2016, Wiczeorek *et al.* 2017, Choi *et al.* 2018, Lee *et al.* 2022), partly validating Remaudière's practical decision. Three additional subfamilies, each including very few species, have been delimited since then, and there are now 23 recognized aphid subfamilies (Favret 2024); but the relationships between these taxa remain unresolved (von Dohlen 2009, Podsiadlowski and Vilcinskis 2016). A few published molecular phylogenetic studies have focused on solving this puzzle. Among the most influential are: the study of Novakova *et al.* (2013) using five DNA markers found on the primary symbiont genome of aphids, *Buchnera aphidicola*; the study of Chen *et al.* (2017) based on the whole mitogenomes from 33 aphid species; and two recent phylogenomic investigations using, respectively, transcriptome data (Owen and Miller 2022) and ultra-conserved elements (Hardy *et al.* 2022), each yielding thousands of nuclear markers (Fig. 1). These studies represent solid attempts at solving the backbone phylogeny of Aphididae in terms of taxa representation, as they include several subfamilies and mostly subfamilies

that encompass the bulk of aphid diversity. They also rely on various and sometimes numerous DNA markers. However, despite these endeavours, the resulting trees exhibit uncertainties in deep nodes. Many relationships between subfamilies lack robust support and depend on the phylogenetic inference methods employed. Additionally, relationships between subfamilies differed between studies with only a few deep nodes shared among all four trees (Fig. 1). Nevertheless, this comparison is inherently challenging due to the lack of overlap in taxa between studies. Consequently, discerning whether the incongruence in datasets arises from differences in taxa sampling, inference methods, or actual differences in phylogenetic signals based on DNA sources remains a complex task.

Here, we aimed at obtaining data from three sources: the mitochondrial genome of aphids, highly conserved nuclear genes from the aphid genome, and genes from the aphid primary symbiont on a set of species representative of aphid diversity. This data was used to: (i) understand whether discrepancies between previously published phylogenies stem from uneven taxa representation and (ii) attempt solving aphid subfamilies relationships. First, we compiled a mitogenome dataset aiming to add subfamily representation to the study of Chen *et al.* (2017) and expand representation of highly diverse subfamilies; to that effect we produced partial or whole mitogenomes that were combined to previously published mitochondria. We also generated multiple independent single-copy nuclear gene sequences, using the Fluidigm Access Array technology and high-throughput sequencing. We then compared the phylogenetic information yielded by these two datasets with the phylogenetic history of *Buchnera aphidicola*. These were obtained from *Buchnera* whole-genome data mainly produced by Chong *et al.* (2019) and Manzano-Marín *et al.* (2023) to which we added 12 new draft genomes to cover additional subfamilies. We tackled how gene partitions obtained from mitochondrial, nuclear, and endosymbiont data align with one another. We further assessed whether the relationships retrieved were congruent with previous views of the evolution of this well-known insect lineage and discuss potential sources of conflicts.

MATERIALS AND METHOD

Taxon sampling

Aphid samples used in this study were sourced from the CBGP INRAe aphid collection (doi: 10.15454/D6XAKL), originally collected from 2006 to 2019 and preserved in 70% alcohol at 6°C. For each sample collected, a preliminary classification of individuals was carried out through observation under the microscope. Subsequently, to confirm the first visual morphological classification, some specimens were mounted on slides and formally identified by A. Coeur d'acier using mainly the keys of Blackman and Eastop (<http://www.aphidsonworldsplants.info/>). After morphological identification under the stereo-microscope, one to 10 specimens, representative of each sample, were randomly selected for DNA extractions and molecular analyses. A full list of samples newly sequenced for this study, voucher numbers as deposited in the CBGP INRAe aphid collection, and sources from previously sequenced aphids are given in the Supporting Information, Table S1. Taxonomy follows Favret (2024).

Mitogenome data

In order to obtain full mitogenomes, we first retrieved reads from the Illumina sequencing of aphid endosymbionts from previous studies (Manzano-Marín *et al.* 2018, 2023, Chong *et al.* 2019, Manzano-Marín *et al.* 2020). They generally yielded aphid mitogenomes with sufficient coverage to allow full assembly. DNA library construction and sequencing protocol are described in respective publications. For both these datasets, reads were trimmed and quality filtered using FASTX-Toolkit v.0.0.14 (<http://hannonlab.cshl.edu/>). Reads shorter than 75 base pairs (bp) were dropped. Additionally, PRINSEQ v.0.20.4 was used to remove reads containing undefined nucleotides, as well as those left without a pair after the filtering and clipping process. The resulting reads were assembled, using SPAdes v.3.10.1 (Bankevich *et al.* 2012), with the options-only-assembler option and k-mer sizes of 33, 55, 77, and 99. Contigs representing aphid mitogenomes were identified based on BlastX searches against a database containing available aphid mitogenomes deposited in

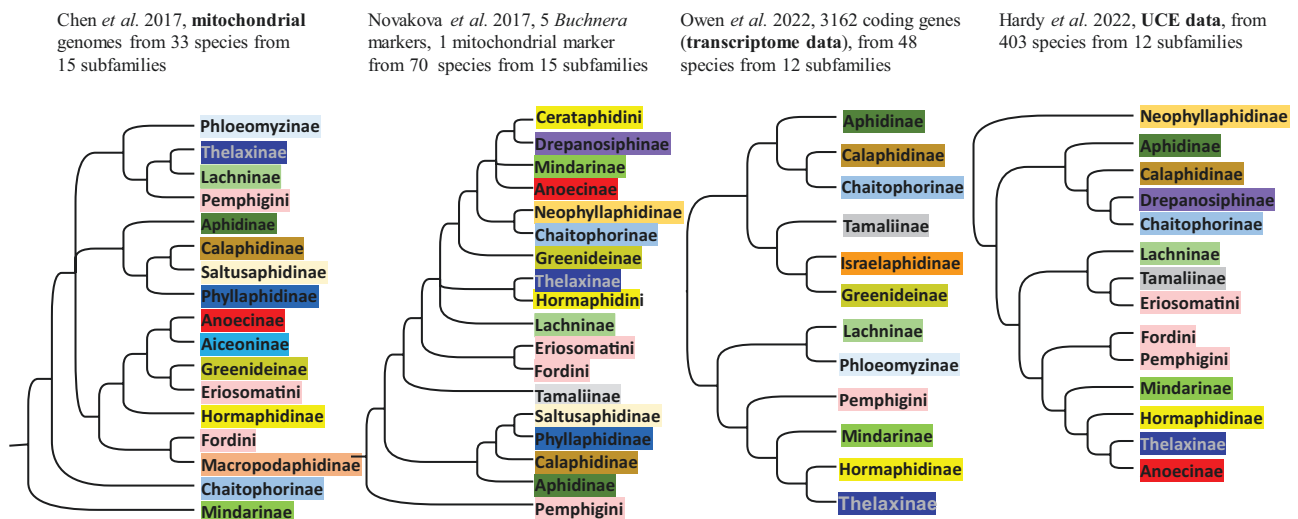


Figure 1. Alternative topologies found in previous phylogenetic investigations of aphids: A, sketch of the Bayesian topology on fig. 1 of Chen *et al.* 2017; B, sketch of the Bayesian topology on fig. 2 of Novakova *et al.* 2013; C, sketch of the ML topology on fig. 1 of Owen *et al.* 2022; D, sketch of the phylogeny obtained by Hardy *et al.* 2022.

the *National Center for Biotechnology Information* (NCBI). The cytochrome *c* oxidase subunit I (*cox1*) genes retrieved in our assemblies were systematically used in BlastN searches against a curated barcode database (based on preparation of specimen vouchers) (Coeur d'acier *et al.* 2014) to confirm species' identification and ensure that we did not have any cross-contamination or chimeric sequences. Mitogenomes generally assembled into a single contig. Through this process, 55 new aphid mitogenomes were assembled: 38 were obtained from endosymbiont sequence dataset from the INRAe-CBGP collection, while 17 were obtained from the endosymbiont sequence dataset of Chong *et al.* (2019) (Supporting Information, Fig S1, Table S1).

In order to cover additional subfamilies, we used polymerase chain reaction (PCR) to amplify three overlapping mitochondrial DNA fragments (long fragments from about 4 kb to 8 kb) on several specimens. For each aphid sample, about 10 to 15 individuals from the same aphid colony were pooled together to obtain DNA extracts of over 10 ng/μL. For DNA extraction, we used the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's recommendations. DNA extractions were then normalized to 10 ng/μL after Qubit fluorimeter quantification (Invitrogen). Long-range PCR amplification and DNA library construction protocols from pooled amplicons and further sequencing are detailed in the Supporting Information, Text S1. Reads' filtering, trimming, assembly, and taxonomic assignments of contigs followed the same procedures as for the endosymbiont sequencing reads. All resulting contigs of the assemblies were identified as parts of an aphid mitogenome, except for a few chimeric contigs or some contigs belonging to aphid parasitoid mitogenomes. Through this long-range PCR procedure, we obtained 28 partial and four complete mitochondrial genomes (Supporting Information, Table S2). For each aphid sample, the contigs were scaffolded against the most closely related aphid mitogenomes available in NCBI or our own assemblies.

All mitogenomes generated were annotated using the MITOS web-server (Bernt *et al.* 2013) and annotations were curated manually using UGENE (10.1093/bioinformatics/bts091) and GENEIOUS v.11.1.5. In brief, we verified start–stop position of protein-coding genes using the annotated reference mitogenomes of Wang *et al.* (2013), and we located and manually annotated a small number of tRNA genes that were missed using blastn searches. Accession numbers for complete and partial mitogenomes obtained in this study are available in the Supporting Information, Table S1. Finally, we completed this dataset with 29 aphid mitogenomes and those of two outgroup species, *Daktulosphaira vitifoliae* and *Adelges tsugae*, available in NCBI. We chose sequences that were complete, preferentially obtained from whole-genome sequencing, in order to detect gene rearrangements if any. Altogether, we gathered 120 mitogenomes.

The gene sequences were aligned using MAFFT v.7.2.2.1 (Katoh and Standley 2013). Alignments for protein-coding genes were corrected to respect translation frames when needed and were translated into amino acid sequences (AA). We produced two data matrices for further phylogenetic analyses: a full DNA matrix concatenating all 13 protein-coding genes (tRNA sequences and ribosomal RNA genes were discarded from analyses as they were ambiguously aligned), and the corresponding AA matrix.

Selection of single-copy nuclear genes, high-throughput PCR on Fluidigm Access Array technology, and high-throughput sequencing

Nuclear gene selection, primer design, and validation

We first targeted nuclear markers that have been previously used in aphid phylogenetic investigations (Elongation factor and long-wave Opsin, Ortiz-Rivas and Martinez-Torres 2010) and then added single-copy genes that have been used in scale insect phylogenetic investigations (Mullen *et al.* 2017). We further targeted single-copy nuclear genes commonly used to solve deep nodes in insect phylogenies (within Lepidoptera; Wahlberg and Wheat 2008, Wahlberg *et al.* 2016) and between insect orders (Wiegmann *et al.* 2009) but never tested on aphids. In order to define appropriate primers for aphids, these genes were searched for in aphid reference genomes available at the time of initiating this study: *Acyrtosiphon pisum* (Acyr 2.0), *Diuraphis noxia* (Dnoxia_1.0), *Rhopalosiphum maidis* (ASM367621v3), *Eriosoma lanigerum* (WAA_JIC_1.0), *Sipha flava* (YSA_version1), and the outgroup *Daktulosphaira vitifoliae* (ASM2509136v1) using either available genome annotations or TblastN. For the latter, the query sequence used for blast searches was the pea aphid sequence of the targeted gene (using min Identity = 80). Genes that had blast hits against all available aphid genomes were then selected along for primer designs using Primer as implemented in GENEIOUS v.11.1.5 (Biomatters, Auckland, New Zealand). For primer design, we followed the recommended criteria specified in the Fluidigm Access Array system protocol (Fluidigm, San Francisco, CA, USA), i.e. annealing temperature was set to 60°C (+/−2°C) for all primers. We targeted amplicons that were 300 to 450 bp long in order to obtain overlapping read pairs through 251bp paired-end sequencing. In order to complete this set of primers we then curated the set of 1478 single-copy orthologous genes listed in Misof *et al.* (2014). Briefly, we first selected nucleotide alignments in which at least two aphid species were present (among the three present in the insect tree, i.e. *Acyrtosiphon pisum*, *Essigella californica*, and *Aphis gossypii*). Among these, we then selected alignments that were longer than 800 bp, in order to work on fragments that were long enough to define conserved primers across aphids. For each of these alignments, the gene sequence from *Ac. pisum* was retrieved and then blasted against aphid reference genomes cited above. Twenty-five genes that had blast hits against all available reference genomes were picked for primer designs, following the same procedure as explained above.

Altogether, a total of 60 pairs of primers were designed. We added Illumina linkers in their 5' position (forward primer: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'; reverse primer: 5'-GTCTCGTGGGCTCGGAGATGTGTAT AAGAGACAG-3') and tested amplification success on eight aphid DNA extracts. We chose samples from species from seven subfamilies to ensure amplification success across the aphid tree. These tests were conducted in 10 μL reaction volume using 5 μL of 2 × QIAGEN Multiplex Kit Master Mix (Qiagen), 0.3 μM of each primer, and 2 μL of DNA extract. The PCR programme consisted in an initial denaturation step at 95°C for 15 min, followed by 40 cycles of denaturation at 94°C for 40 s, annealing at 55°C for 45 s, and extension at 72°C for 60 s, followed by a final extension step at 72°C for 10 min. Forty-seven primer pairs that amplified all seven DNA extracts were validated. A complete

list of the chosen primers as well as gene names are found in the [Supporting Information, Table S3](#).

Microfluidic amplification and sequencing

We amplified these 47 markers on 96 aphid species trying to maximize overlap with species represented in the mitogenome data. We used the DNA extracts used for mitogenome amplification to which we added representatives of aphid species from additional subfamilies and tribes ([Supporting Information, Table S1](#)); DNA extraction protocols were conducted as for aphid material produced for long-range PCRs.

Microfluidic PCRs were performed in an Access Array System (Fluidigm) using two fully loaded 48 × 48 Access Array Integrated Fluidic Circuits (Fluidigm): this allows for 48 samples to be simultaneously amplified across 48 distinct primer pairs. We amplified on each array the same 47 genes on 48 samples, leaving a blank template with water and no primer pair as a negative control on one of the arrays. We followed the manufacturer's protocols, except for the PCR conditions for which we used the QIAGEN Multiplex Kit Master Mix (Qiagen) and the same PCR programme as for the primer tests (hybridation 30°C and 50°C) but preceded by incubation at 50°C for 2 min and 70°C for 20 min. The PCR products were then harvested from each array for each sample (each aphid species) in 96 pools of 10 to 12 µL. Eleven pools were verified on a BioAnalyzer High Sensitivity DNA Assay (Agilent Technologies, Santa Clara, California, USA). The 96 samples were PCR-indexed using the UDI (Unique Dual Indexes) from [Martin \(2019\)](#) and the procedure described in [Galan et al. \(2018\)](#). All 96 samples were then pooled together in equal volume and 83 µL of this final pool was subjected to size selection for the full-length amplicons (expected size >350bp including primers, indexes, and adaptors) by excision on a low-melting agarose gel (1.25%). This enabled to discard non-specific PCR products and primer dimers. The PCR Clean-up Gel Extraction kit (Macherey-Nagel) was used to purify the excised PCR products. It was then analysed in a Bioanalyzer High Sensitivity Assay (Agilent Technologies, Santa Clara, California, USA) and standardized at 4nM after a KAPA qPCR quantification (KK4835; Kapa Biosystems, Woburn, Massachusetts, USA). The pooled amplicons were then submitted for paired-end sequencing on a MiSeq (Illumina) flowcell using a 500-cycle reagent cartridge v.3.

Bioinformatic treatment

After filtering through the Illumina's control quality procedure, reads were demultiplexed for each aphid sample, using the sample-specific dual barcode combinations and then, for each nuclear locus, using the target specific primers. We then used a custom script from [Sow et al. \(2019\)](#) to merge paired sequences into contigs with FLASH v.1.2.11 and trim primers with cutadapt v.1.9.1. Merged reads for each sample and each locus were then imported into GENEIOUS v.11.1.5. Loci that were represented by less than 25 reads were discarded from the dataset. For each locus, each aphid sample reads were first dereplicated then assembled using GENEIOUS assembler (using options: high sensitivity, saving 10 contigs and generating consensus for each contig with a 85% identity threshold). The assemblies generally produced a single contig per locus per aphid sample. On a few

occasions, when several contigs were obtained, one had a much higher coverage (1000 times higher than the other) and the low coverage contigs were discarded.

Gene sequences were aligned using MAFFT v.7.2.2.1 ([Katoh and Standley 2013](#)). Alignments were corrected to respect translation frames. Some of the genes contained intergenic regions, which were difficult to align between samples belonging to different subfamilies. The intergenic regions were, therefore, deleted from the final nucleic acid dataset. Finally, protein-coding genes were translated into AA sequences. We concatenated all nucleic acid sequences to obtain a DNA matrix and all translated sequences to obtain an AA matrix.

Buchnera genome data

We used 61 published genome sequences, mainly from [Chong et al. \(2019\)](#) (16 genomes) and [Manzano-Marin et al. \(2018, 2020, 2023\)](#) (29 genomes), for which taxon sampling overlapped with the one of aphid mitochondrial genomes. *Escherichia coli* K-12 MG 1655 and a selection of *Pantoea* and *Eriwnia* strains were used as outgroups. In addition, we sequenced 12 new *Buchnera aphidicola* genomes ([Supporting Information, Table S1](#)). For both sequencing and assembly of this new dataset, we followed the protocols of [Manzano-Marin et al. \(2023\)](#). Draft genomes were made of one to nine contigs ([Supporting Information, Table S4](#)) representing potentially incomplete *Buchnera*. Coding sequences (CDS) were searched in these draft *Buchnera* genomes using PRODIGAL 2.6.3 ([Hyatt et al. 2010](#)).

The search for orthologue groups was carried with OrthoFinder-2.5.4 ([Emms and Kelly 2019](#)) using as input Prodigal's AA results for newly sequenced *Buchnera* and AA sequences from annotated published genomes. We then retrieved the single-copy core proteins of the selected genomes for phylogenetic reconstruction. We aligned the single-copy core protein sets, gene by gene, using MAFFT v.7.2.2.1 ([Katoh and Standley 2013](#)) (L-INS-i algorithm). Divergent and ambiguously aligned blocks were removed using GBlocks v.0.91b ([Talavera and Castresana 2007](#)). The resulting alignments were concatenated for phylogenetic inference.

Phylogenetic analyses

The workflow of phylogenetic analyses applied on the three datasets is summarized in the [Supporting Information, Figure S1](#).

For each aphid gene partition (mitogenomes and nuclear genes), we analysed the DNA and AA supermatrices resulting from the concatenation of all genes and implemented several models of substitution using partitioned and non-partitioned approaches. For non-partitioned analyses, we chose the best model using ModelFinder as implemented in IQ-TREE v.2.1.3 (option MF) ([Nguyen et al. 2014](#)). For partitioned analyses, to choose the most appropriate partitioning schemes and best models, for each matrix we built a pre-partition file: for DNA data we divided each protein-coding sequence into three partitions, one for each codon position; for AA data each gene was delimited. We then ran ModelFinder (IQ-TREE v.2.1.3 option MFP + MERGE), which both tests models for each partition and tests if partitions can be merged. Once models and partitions were selected, we ran maximum likelihood (ML) analyses

in IQ-TREE v.2.1.3 with 1000 ultrafast bootstrap replicates (-bb option), using the best-fitting models and partitions found based on BIC (Bayesian information criterion) model selection.

We also used a class of models called profile mixture models in which the likelihood of the data at each site in the alignments is calculated as a weighted sum across multiple classes, each class having its own substitution rate matrix (Lartillot *et al.* 2007). Modelling site-heterogeneity patterns has often been shown to outperform partition strategies in recovering tree topology (Wang *et al.* 2019). It can mitigate LBA (long-branch attraction) artefacts (Lartillot *et al.* 2007) and recover contentious relationships, such as relationships of bacterial endosymbionts with free-living bacteria (Husnik *et al.* 2011) or the branching of bacteria within Alphaproteobacteria (Fan *et al.* 2020). Such a model is implemented in PHYLOBAYES for Bayesian approaches (Lartillot *et al.* 2009) and in IQ-TREE v.2.1.3 (Nguyen *et al.* 2014) under ML optimization. Using PHYLOBAYES 4.1C, we first ran the CAT model (Lartillot and Philippe 2004) on all AA matrices. For each dataset, we ran two independent Markov chains Monte Carlo (MCMC) analyses, starting from a random tree until the two chains converged [i.e. effective population size >50 for all parameters and maximum discrepancy (maxdiff) of 0.3]. Sampling trees and associated model parameters every cycle, we then computed a consensus tree with the posterior probability (PP) of each node. As stated above, IQ-TREE v.2.1.3 provides a number of site-specific frequency models and profile mixture models for AA sequences; for instance, C10 to C60 a variant of the CAT model applied in PHYLOBAYES. In order to select the best model in ML analyses, we ran the -MF option on non-partitioned AA matrices, including several of these mixture models (C10, C20, and C50) with the -madd option in IQtree v.2.1.3 and then ran an ML analysis using the best-fitting models and 1000 ultrafast bootstrap replicates. On the DNA matrices, for each dataset (mitochondrial and nuclear), we implemented a mixture of the models selected previously on the partitioned matrices using the -m mix option. We then ran ML analyses in IQ-TREE v.2.1.3 with 1000 ultra-fast bootstrap. [Table S5 in the Supporting Information](#) lists the models implemented for each dataset under ML analyses and corresponding likelihoods of the best trees.

Contrary to the mitochondrial DNA, which is inherited as a single locus, the nuclear genes might show genealogical discordance because of incomplete lineage sorting (ILS). In order to analyse whether the lack of signal came from conflicting loci or an overall lack of information, we measured the fraction of loci and the fraction of sites consistent with a particular branch inferred in our analyses of the supermatrix (Minh *et al.* 2020). Using the ML tree obtained above [Figure 3](#) as a reference tree, for each node, the gene concordance factor (gCF), i.e. the proportions of inferred single-locus gene trees showing that node, were estimated. We used the options implemented in the IQ-TREE software package that takes into account variation in taxon coverage of each gene (Minh *et al.* 2020). We then estimated site concordance factors (sCF), a measure that estimates concordance at the level of sites. These node support values complement traditional bootstraps and PP values, and help to decipher the cause of topological variation in phylogenetic datasets. These analyses

were conducted on the nucleotide dataset as the individual gene sequences in AA were often too short to reconstruct a gene tree.

For *Buchnera* data analyses, given the high evolutionary rate of endosymbiont genomes (Moran *et al.* 1995, Wernegreen *et al.* 2001), only AA alignments were used for further analysis. The selection of substitution models on the datasets was carried out using ModelFinder (Kalyaanamoorthy *et al.* 2017) (option -MFP to test classical protein evolution models in IQ-TREE v.2.1.3) using the BIC. We further performed ML searches using the selected model and an additional mixture model, C20. Bayesian inferences were conducted with the CAT model (Lartillot and Philippe 2004), as implemented in PHYLOBAYES 4.1c (Lartillot *et al.* 2009), as explained for nuclear and mitochondrial datasets.

Looking for sources of conflicting topologies

In order to better compare the topologies obtained with different DNA sources, we selected only species for which we obtained data from all three sources. We, therefore, worked on a subset of 51 aphid species that were represented by mitogenome data, nuclear genes, and *Buchnera* genome data; except in two cases (namely *Rhopalosiphum* spp. and *Cavariella* spp.) in which, for the *Buchnera* data, we selected species from the same genus. We first measured statistical supports for alternative topologies using the approximately unbiased test (AU test) (Shimodaira 2002). We tested whether the AA matrix of the mitogenomes rejected or not alternative trees, i.e. nuclear and *Buchnera* ML trees obtained with C50 model for each respective dataset. In order to do that, trees were modified to enforce subfamily relationships observed in alternative topologies, with branch lengths all set to 1. Tests were performed on the IQ-TREE v.2.1.3 with 10 000 replicates.

We then concatenated aphid nuclear and mitochondrial AA matrices in a single supermatrix and ran a new ML search using a protein mixture model (C50) implemented in IQ-TREE v.2.1.3. We ran a ML analysis on a *Buchnera* dataset with the same 51 species, using also C50 substitution model. Both trees were visually compared.

We further explored whether discrepancies between aphid and *Buchnera* topology could stem from LBA effects in the *Buchnera* topology. Indeed, recent studies showed that several aphid subfamilies host very small *Buchnera*. These reduced genomes are often associated with the occurrence co-obligate symbioses (Manzano-Marín *et al.* 2023), i.e. the association of *Buchnera* with a secondary nutritional symbiont that complements some essential functions lost by the primary symbiont. This dual symbiotic system has been hypothesized to accelerate genome erosion in *Buchnera* (Lamelas *et al.* 2011) and could lead to convergent evolution in *Buchnera* genome, generating artefacts in the phylogeny. In order to explore this potential cause of conflicting phylogenies between *Buchnera* and its hosts, we investigated evolutionary rate shifts along branches of the *Buchnera* phylogeny. As a first test of signatures of convergent molecular evolution, we asked whether branches of the *Buchnera* tree where dual-symbiotic had evolved, underwent different rates of genome-wide molecular evolution compared to 'mono-symbiotic branches'. We worked on a dataset limited to species where the absence/presence of dual-symbiotic systems had been investigated (Mesguier *et al.* 2017, Manzano-Marín *et al.* 2018,

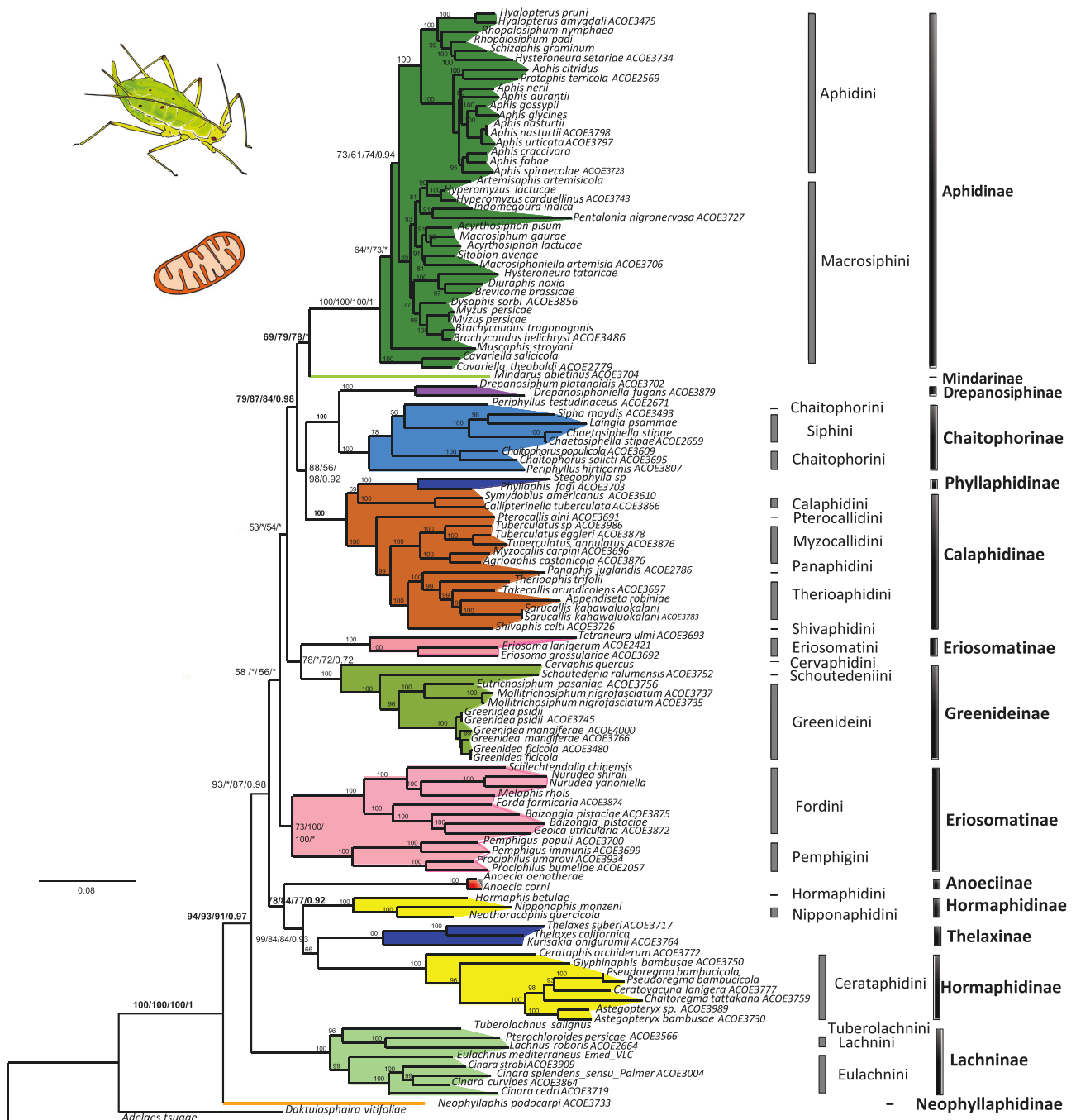


Figure 2. Trees obtained from the analysis of whole mitogenomes. Best tree found with ML analyses of the AA matrix under the best-fitting model (mMet+F+I+R6), values at nodes indicate bootstrap values obtained with best-fitting model/the C50 model/partitioned analyses/ and posterior probabilities under the CAT model of PhyloBayes; when a "*" is indicated instead of a support values it means that the nodes were not retrieved in the corresponding analysis or that support was below 50. Alternative support values are only given for deep nodes that were inconsistent across analyses.

2023, Manzano-Marín *et al.* 2020, Monnin *et al.* 1949, Renoz *et al.* 2022), i.e. a subset of 60 *Buchnera* genomes. We estimated the genome-wide ratio of non-synonymous/synonymous substitution (i.e. dN/dS) across the phylogeny reduced to this dataset, those were inferred using the free-ratios model of PAML v.4.9e (Yang 2007).

We then used the relative rate test implemented in the R package RERconverge (Kowalczyk *et al.* 2019) (<https://github.com/nclark-lab/RERconverge>)

to look for convergent changes in evolutionary rates in *Buchnera* genes in dual symbiotic systems. Using a species' phylogenetic tree, RER converge calculates relative branch lengths for a specific gene by normalizing branch length to the distribution of branch lengths across all genes. Genes for which these standardized rates are consistently either higher or lower in foreground branches (i.e. defined as branches leading to dual symbiotic lineages) compared to

background branches (i.e. all remaining branches) are identified as having experienced accelerated or decelerated molecular evolution in dual symbiotic systems. The analysis was conducted on the same set of 146 genes used for phylogenetic reconstruction. We tested for significant differences in rates within the 13 di-symbiotic lineages relative to the rest of the phylogeny. The reference phylogeny used for this analysis was the ML topology.

RESULTS

Mitogenome phylogeny

Altogether, we assembled in this study 55 new complete and 32 partial mitochondrial genomes. The long-range PCR procedures generally produced one to three contigs that overlapped but did not fully recover the entire mitogenomes (Supporting Information, Table S2). This stemmed from failures to amplify some long-range fragments. The fragment covering *COI* to *COII* was generally easily retrieved, while the two other targeted fragments were amplified with further difficulties. Looking at the dataset, including all complete mitogenomes, we observed very few genome rearrangements, with only a shift in the position of the gene cluster *trnF-nad5-trnH-nad4-nad4l* from being flanked on its 3'-end by *trnP-nad6-cob-trnS2* to being flanked by the same gene cluster on its 5'-end as observed by Zhang *et al.* (2021) on one species of Lachninae. This shift is shared by the 13 Lachninae and the two Thelaxinae samples for which we obtained complete mitochondrial genomes. After concatenation of the 13 coding genes of the 88 new mitogenomes and 31 published mitogenomes (Supporting Information, Fig. S1), we obtained a DNA supermatrix of 120 sequences of 11 407 bp. The total AA matrix after concatenation of the 13 translated genes, was 3782 AA long.

Results of ML analyses of the concatenated AA matrix are presented in Figure 2 with support values obtained across alternative ML analyses and PHYLOBAYES analyses (alternative AA topologies and topologies obtained with nucleotidic sequences are available in the Supporting Information, Archive S1). In all analyses, most subfamilies were recovered as monophyletic, whether AA or nucleic acid sequences were analysed and across substitution models. The exceptions are Eriosomatinae samples that were either scattered into three to two clades and the Calaphidinae specimens that systematically included Phyllaphidinae representatives. Depending on substitution models used for the AA matrix, Thelaxinae species either clustered within Hormaphidinae or, alternatively, were found as a sister-group to all Hormaphidinae (PHYLOBAYES and ML analyses under the C50 model). Tribes represented in our dataset appeared monophyletic in all analyses, except for Chaitophorini and Macrosiphini. In the latter, *Cavariella* spp. and *Muscaphis stroyani* (Smith, 1980) appeared as sister-groups to all Aphidinae and were not included within Macrosiphini. Many subfamily relationships were characterized by short branches with low support values; they were also highly sensitive to substitution model specification (i.e. not retrieved in all analyses). The only subfamily relationships that were strongly sustained and consistent across analyses were: (i) the position of Neophyllaphidinae as a sister-lineage to all other aphids; (ii) the basal position of Lachninae; (iii) the clustering of Chaitophorinae with Drepanosiphinae; (iv) the clustering of Calaphidinae and Phyllaphidinae; and (v) the clustering of Hormaphidinae with Thelaxinae and Anoeciinae. Additionally,

a group made of Calaphidinae (including Phyllaphidinae), Chaitophorinae, and Drepanosiphinae, consistently clustered with Aphidinae but with relatively low support values (bootstrap < 90).

Nuclear gene phylogeny

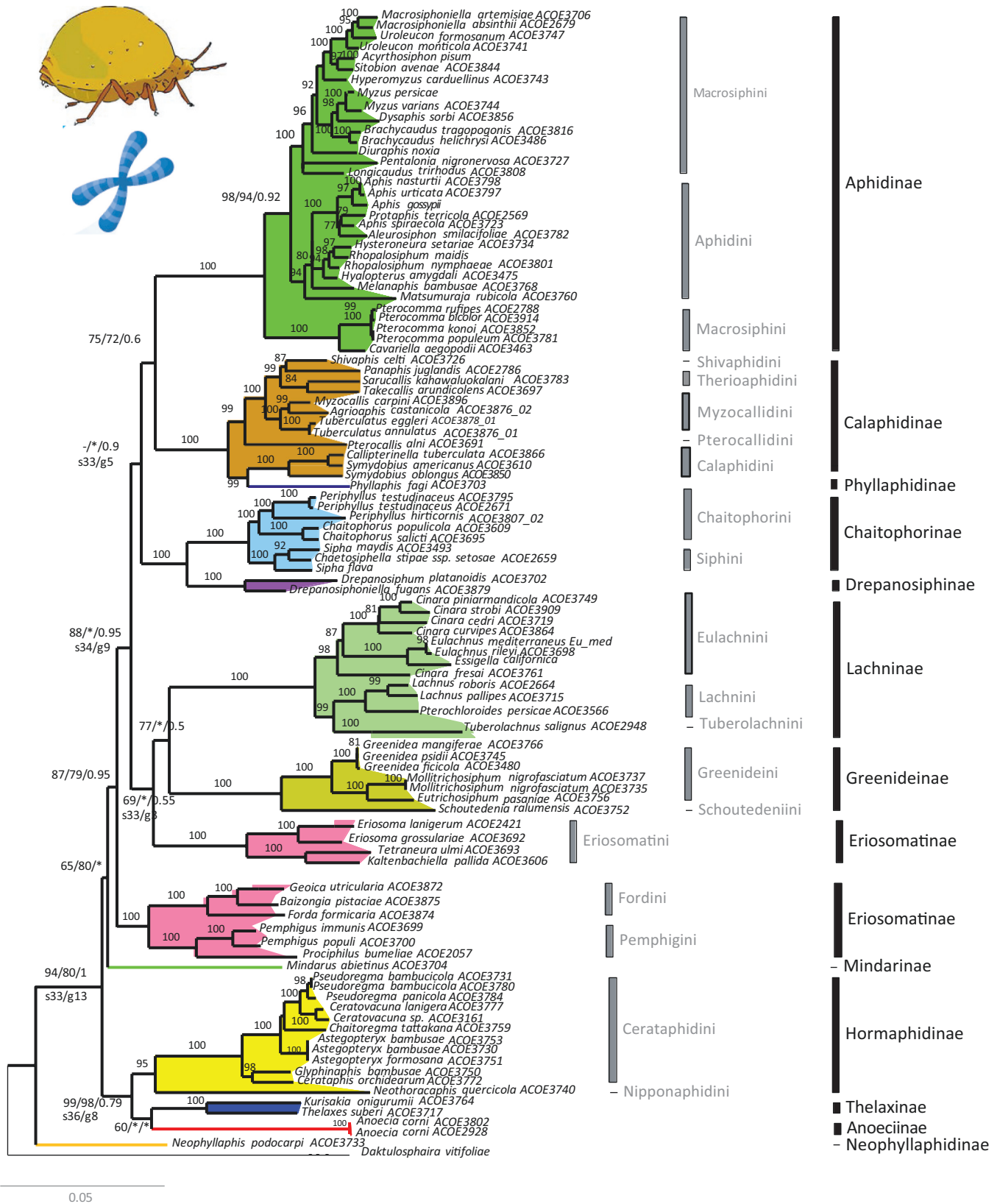
Altogether, we obtained the sequences of 3146 amplicons across 101 aphid specimens representing 99 aphid species. We failed to obtain contigs with sufficient coverage for five of the targeted loci; those were further excluded from the dataset. Sequencing success for each amplicon is given in the Supporting Information, Table S3 and number of loci per species is given in the Supporting Information, Table S1. After concatenation of the 42 loci, we obtained a DNA supermatrix of 13 400 bp, 10 504 bp once intergenic regions were excluded. The total AA matrix after concatenation of the 42 translated genes was 3486 AA long. Results are presented in Figure 3 with support values obtained across alternative ML analyses and PHYLOBAYES results, as well as gene and site concordance factors obtained on the DNA matrix (alternative AA topologies and topologies obtained with nucleotidic sequences are available in the Supporting Information, Archive S2).

Phylogenetic analyses confirmed the monophyly of most subfamilies, with the exception of Calaphidinae, that included Phyllaphidinae and Eriosomatinae. The latter was divided into two clusters: one made of the Eriosomatini tribe and another made of the Pemphigini and Fordini tribes. Tribes were mostly monophyletic, with the exception of Chaitophorini and Macrosiphini as for the mitochondrial phylogeny. Paralleling the mitochondrial trees: (i) Neophyllaphidinae appeared as a sister-lineage of all aphids; (ii) Chaitophorinae clustered with Drepanosiphinae; (iii) Phyllaphidinae was positioned within Calaphidinae; and (iv) Hormaphidinae clustered with Thelaxinae and Anoeciinae, but their relative positions within that group were not similar across analyses (Fig. 3). Chaitophorinae, Drepanosiphinae, Calaphidinae (including Phyllaphidinae), and Aphidinae formed a clade in ML analyses of the AA matrix under the partitioned analysis and the C50 model, but this relationship had very low support and was not retrieved in ML analysis using the Q_{plant}+F+I+R4 model, where Calaphidinae (including Phyllaphidinae) was quite basal. However, this clade was retrieved in all analyses of the DNA matrix (Supporting Information, Archive S2). The remaining subfamily relationships did not converge with the mitochondrial phylogeny and were unstable across analyses, exhibiting very short branches.

The concordance factor analyses conducted on the DNA dataset revealed that short, internal branches that generally conflicted between analyses were retrieved in very few genes (low gCF, i.e. at the most 13 of the 42 genes) and were generally supported by a sCF of 33%. The latter value is what is expected when there is no consistent signal in an alignment (Minh *et al.* 2020) (sCF are calculated for quartets, so a single site can only support one out of three topologies). The low gCf values were, therefore, probably due to a very weak signal in each locus and a not strong gene discordance caused, for instance, by ILS.

Buchnera phylogenetic hypothesis

We retrieved 147 highly conserved shared genes between the 73 bacterial genomes included in our analyses (i.e. 69 *Buchnera* and



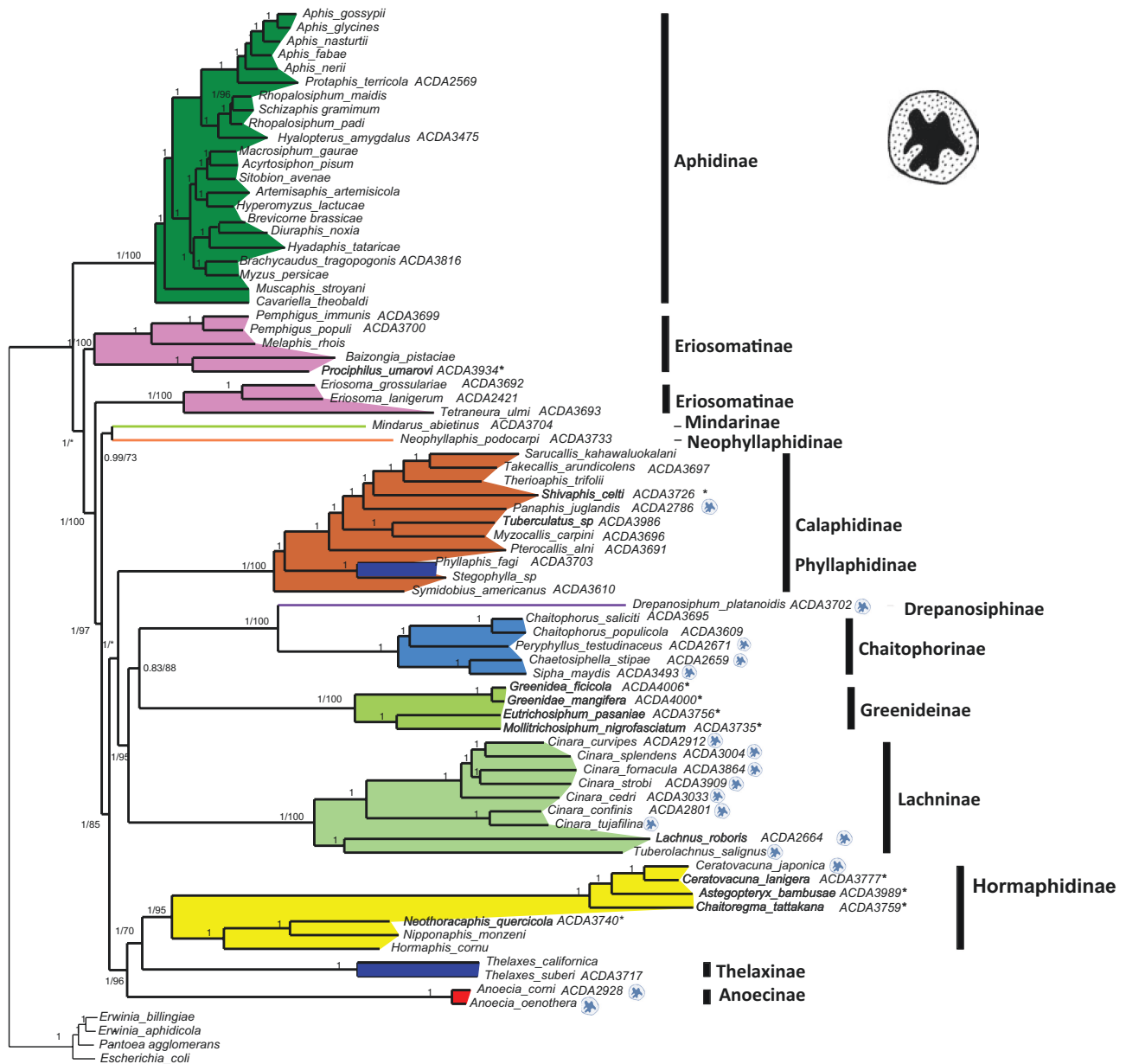



Figure 4. *Buchnera* tree obtained from the analysis of 146 orthologous shared single-copy genes. The topology corresponds to the consensus tree found under Bayesian analyses with the CAT model of PHYLOBAYES: values at nodes indicate posterior probabilities in Bayesian analyses/ultrafast bootstrap values under ML analysis with the C20 model/ultrafast bootstrap values under ML analysis with the cpREV+F+I+R8 model. * at nodes indicate bifurcations that were not observed in the ML analysis.  following species names indicate species for which dual-obligate symbiosis has been demonstrated, * following species names indicate species for which the symbiotic status (mono-symbiotic *Buchnera* or disymbiotic system) is unknown. Names in bold are species for which new *Buchnera* data has been acquired for this study.

four outgroups). This low number in comparison to [Manzano-Marín *et al.* \(2023\)](#) (retrieving 229 genes) probably stems from the inclusion of 12 incomplete *Buchnera* draft genomes mainly from Greenideinae ([Supporting Information, Table S4, Archive S3](#)). After concatenation and deletion of ambiguously aligned positions, we obtained a matrix of 33 820 AA. MF selected cpREV+F+I+R8 as the best-fitting model. The *Buchnera* phylogeny obtained with this dataset appeared well resolved with ML and PHYLOBAYES analyses converging towards a similar topology sustained by strong bootstrap and PP values ([Fig. 4](#)). As in the aphid datasets, subfamilies appeared monophyletic,

except for Eriosomatinae that were scattered in three or two clusters depending on analyses (i.e. ML vs. Bayesian analyses; [Supporting Information, Archive S4](#)) and Calaphidinae that systematically included Phyllaphidinae. As in trees based on aphid data, Chaitophorinae clustered with Drepanosiphinae, while Hormaphidinae always clustered with Thelaxinae and Anoeciinae. Other relationships between subfamilies were in strong disagreement with inferences from aphid genome partitions, but they were generally robust and consistent across analyses. They generally positioned the Aphidinae subfamily (as well as Pemphigini in ML searches, [Supporting Material,](#)

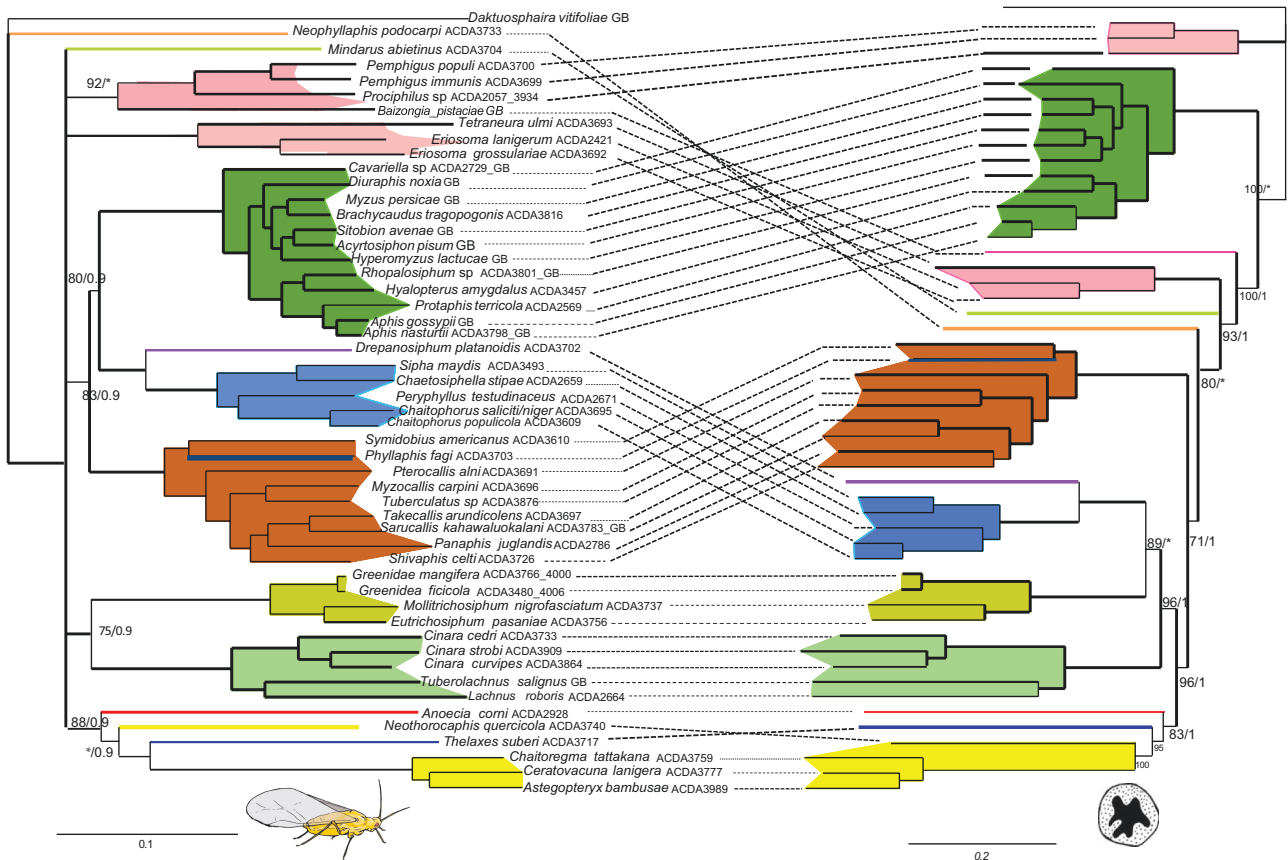


Figure 5. Coplot of the combined analysis of mitochondrial and nuclear data vs. *Buchnera* ML tree under C50 mode, I for species that have data from all three sources, ultrafast bootstrap values are given for subfamily relationships and PP values obtained in PHYLOBAYES analyses of the same dataset. Branches with support below 50 have been collapsed.

Archive S3) as a sister-group to all other aphids. The clustering of Chaitophorinae and Drepanosiphinae, with Greeneidinae and Lachninae was consistent across analyses but the relative position of Calaphidinae (including Phyllaphidinae), as well as the position of Pemphigini, varied (Fig. 4).

Analysing congruence and conflicts within and between datasets

When applying a congruence test (AU test in IQtree) to the mitogenome dataset, we found that the nuclear tree was not rejected ($P = 0.247$), while the *Buchnera* topology was rejected ($P = 0.002$). We then analysed a concatenated aphid data matrix (mitochondrial and nuclear) restricted to a subset of 51 aphid species that were represented by mitogenome data, nuclear genes and *Buchnera* genes.

The tree obtained from the concatenation of aphid mitochondrial and nuclear genes was very similar to the tree obtained with nuclear data only, but retrieved the clustering of Aphidinae with Calaphidinae/Phyllaphidinae and Chaitophorinae/Drepanosiphinae with good support. When compared to the *Buchnera* tree obtained on the same subset of taxa, we observed a perfect co-speciation pattern between *Buchnera* and its aphid hosts within subfamilies (or tribes in the case of Eriosomatinae) (Fig. 5). On the other hand, relationships between subfamilies in the aphid and *Buchnera* tree are in disagreement, except for the clustering of Thelaxinae, Hormaphidinae, and Anoeciinae,

the clustering of Calaphidinae and Phyllaphidinae, and the clustering of Chaitophorinae and Drepanosiphinae (Table 1).

Analyses of molecular evolution patterns in *Buchnera* revealed no significant differences in overall dN/dS ratio (Supporting Information, Fig. S2) between branches leading to dual-symbiotic lineages and branches leading to monosymbiotic lineages (Wilcoxon rank test, $W = 429$ $P = 0.3045$). Out of 138 single-copy orthologues included in the RERconverge analysis, we identified eight genes that were potentially under convergent acceleration or deceleration in dual symbiotic systems (Supporting Information, Table S6). Our aim here was to test whether such genes could bias the results of phylogenetic analyses; we thus deleted them from the global matrix and reran an ML analysis under the C50 model. We obtained a phylogenetic tree that was identical to the tree including these genes, suggesting that they do not generate LBA in the *Buchnera* tree (Supporting Information, Archive S4).

DISCUSSION

Resolving the phylogeny of Aphididae is the scaffolding for a broader understanding of their evolutionary history. Previous attempts to construct a robust phylogenetic hypothesis for this family have yielded only partially satisfactory results. Mitogenomes have resolved many recalcitrant relationships across insects (e.g. Cameron *et al.* 2007, Łukasik *et al.* 2019,

Table 1. Previous evidence of the strongly sustained relationships found in our study, based on molecular studies and morphological investigations. We are knowingly not exhaustive in our review of taxonomic classifications as those have been numerous (see review of [Wojciechowski 1992](#)). *In the classification of Heie, 1980, the subfamily Drepanosiphinae includes current representants of Drepanosiphinae, Chaitophorinae, Calaphidinae and Phyllaphidinae but also many other representants of small subfamilies (e.g. Neophyllaphidinae, Spicaphidinae but not exclusively).

Subfamily relationships	Phylogenetic evidence from this study	Previous molecular phylogenies retrieving this clustering	Classifications clustering genera that are now forming these subfamilies	Shared morphological characters (not discussed in Ortiz-Rivaz <i>et al.</i> 2010)
Theanaxinae + Hormaphidinae + Anoeciinae	<i>Buchnera</i> + mitochondria + nuclear genes	Ortiz-Rivas <i>et al.</i> 2004 , Ortiz-Rivas and Martinez-Torres 2010 , Hardy <i>et al.</i> 2022	Börner 1952 , Borner and Heinze 1957 (the latter also included current Mindarinae and Phloemyzinae)	Apterae with head capsule and pronotum completely or nearly (in Anoeciinae) fused, dwarfish males with modified genitalia (e.g. fused testes), wingless sexuals with developed mouthparts
Chaitophorinae + Drepanosiphinae	<i>Buchnera</i> + mitochondria + nuclear genes	Ortiz-Rivas <i>et al.</i> 2004 , Ortiz-Rivas and Martinez-Torres 2010 , Wieczorek <i>et al.</i> 2017 , Hardy <i>et al.</i> 2022 .	Baker 1920 , Borner 1930 , Heie 1980*	No sclerotisation of segment II of the rostrum, normal sized males, genitalia not modified
Calaphidinae + Phyllaphidinae	<i>Buchnera</i> + mitochondria + nuclear genes	Novakova <i>et al.</i> 2013 , Chen <i>et al.</i> 2017 , Lee <i>et al.</i> 2022	Börner 1952 , Borner & Heinze 1957 , Bodenheimer & Swirski, 1957 , Heie 1980*	Normal sized males, genitalia not modified
Aphidinae + (Chaitophorinae + Drepanosiphinae) + (Calaphidinae + Phyllaphidinae)	mitochondria + nuclear genes	—	Borner, 1930 , Mackauer, 1965 , Börner 1952 , Wojciechowski 1992 (the two latter classifications also included Lachninae)	Normal sized males, genitalia not modified

[Chang *et al.* 2020](#)). A previous study had suggested that it could be the case for Aphididae ([Chen *et al.* 2017](#)). Nevertheless, our analyses based on a larger sampling reveal that many internal branches remain unresolved and bear a limited amount of phylogenetic signal. Similarly, our set of nuclear genes provides restricted resolution at deeper evolutionary scales. Notably, the topologies derived from nuclear and mitochondrial datasets exhibit some incongruence. The phylogeny obtained with *Buchnera*, which genome has recurrently been used as a source of data for resolving evolutionary relationships in aphids ([Jousselin *et al.* 2009](#), [Novakova *et al.* 2013](#)), further yields a discordant history. Although this latter phylogenetic hypothesis appears more robust than those derived from aphid DNA datasets, strong discordance between this hypothesis and previous views on the evolution of aphids raises questions about its accuracy and, more generally, whether a singular tree that best approximates the evolutionary relationships within Aphididae can be attained. We first discuss phylogenetic relationships retrieved across datasets in light of the taxonomic literature and subsequently explore sources of conflicts between datasets.

Subfamily monophyly and subfamily relationships across datasets

Across all datasets, subfamilies as defined by [Remaudière and Remaudière \(1997\)](#) appear monophyletic, except for

Eriosomatinae, which are generally scattered into two groups (a cluster made of Pemphigini and Fordini and a cluster made of Eriosomatini) and Calaphidinae, which always include Phyllaphidinae. The non-monophyly of Eriosomatinae was found in previous phylogenetic studies of Aphididae ([Novakova *et al.* 2013](#), [Hardy *et al.* 2022](#)) and studies dedicated to the subfamily [e.g. [Qiao and Zhang \(2008\)](#), where Eriosomatinae are referred to as Pemphiginae]. Aphids assigned to Eriosomatinae have long been set aside from other aphid groups: they share the particularity of having dwarf sexual morphs with degenerate mouthparts ([Heie 1980](#)). The unstable position of the tribes across reconstructions makes it difficult to infer the evolutionary trajectory of this characteristic, but phylogenetic reconstructions obtained so far suggest that it might have evolved several times. Eriosomatinae share a similar type of host-alternation (i.e. the occurrence of a generation of sexuparae) and the ability to form galls with Hormaphidinae, they therefore, have often been suggested to be closely related to this subfamily ([Ghosh 1985](#)). But all analyses presented here reject this hypothesis, and imply that both host-alternation and gall-making have evolved independently in Hormaphidinae and Eriosomatinae, and maybe even several times within Eriosomatinae. The paraphyly of Calaphidinae, which in our study include Phyllaphidinae, was retrieved in all our analyses and contrasts with the results obtained by [Lee *et al.* \(2022\)](#). This latter study, focusing on resolving the phylogenetic history of Calaphidinae, placed Phyllaphidinae as a sister-taxon

of Calaphidinae, though the latter included representatives of Saltusaphidinae (not represented in our study), also rendering the subfamily paraphyletic. Though our phylogeny is based on more markers, it encompasses less species, which might bias the relative position of Phyllaphidinae and Calaphidinae is consistent across studies. Because, Phyllaphidinae and Calaphidinae delimitations have varied throughout classifications, it is hard to investigate how these findings align with early views on the evolution of aphids. Our results nevertheless seem to mirror the taxonomic discussions that resulted in including ‘Phyllaphidinae’ in a group called ‘Callaphididae’ [see classifications of Börner (1952) and, Lee *et al.* (2022)] (Table 1).

The proximity of subfamilies Chaitophorinae and Drepanosiphinae is consistently found across our analyses. This agrees with previous aphid molecular phylogenies (Ortiz-Rivas *et al.* 2004, Ortiz-Rivas and Martinez-Torres 2010, Hardy *et al.* 2022), and more specific phylogenetic investigations within Chaitophorinae (Wieczorek *et al.* 2017) (Table 1). The validity of this clustering in relation to morphological investigations has been thoroughly discussed in Wieczorek *et al.* (2017). The affinity of both subfamilies is supported by the similarity of their parasitoids (Mackauer 1965) and morphological data (Table 1) (Wojciechowski and Wieczorek 2004, Wieczorek *et al.* 2011, 2017).

We also systematically recovered sister-group relationships between Anoeciinae, Thelaxinae, and Hormaphidinae. This again agrees with early investigations and some recent phylogenetic studies (Table 1). These three aphid groups were initially clustered in Börner’s classification within a group that also included Mindarinae and Phloeomyzinae. In addition to the characters underlined by Ortiz-Rivas *et al.* (2004), apterae of all these subfamilies share a head capsule and pronotum fused with reduced (i.e. not compound) eyes (Börner 1952). They also share dwarfish males (i.e. less than half the size of the female) with modified genitalia, no wings, and developed mouthparts (Wieczorek *et al.* 2011, 2012) (Table 1). This latter character differentiates them from Eriosomatinae males, which have degenerated mouthparts (Wieczorek *et al.* 2011). Hormaphidinae and Thelaxinae winged morphs also have the particularity of having wings that lie flat on the body, unlike the roof-like position common among other aphids (Heie 1980, Miyazaki 1987). Hence, in light of their morphology and early taxonomic literature, the phylogenetic clustering of these three subfamilies is not surprising.

All other subfamily relationships are inconsistent across analyses. In particular, the relative positions of the two species-poor subfamilies Neophyllaphidinae (18 species) and Mindarinae (nine species), both represented by a single specimen each in our study, are variable. They have been hypothesized to be sister-groups of other aphids based on their association with gymnosperms (Russell 1982, Heie 1985) and their overall morphological similarities with aphid-like fossils from the Upper Triassic (Heie 1967, Szwedo and Nel 2011). In particular, *Neophyllaphis* aphids closely resemble Cretaceous *Aniferella* fossils (von Dohlen 2004). This hypothesis finds some support in our aphid data analyses for Neophyllaphidinae as *Neophyllaphis podocarpus* (Takahashi, 1920) appears as a sister-taxon to all

other aphids. This is also the case in the topology retrieved by Hardy *et al.* (2022). However, this basal position is not sustained in the *Buchnera* tree. Concerning Mindarinae, their position is more variable across trees. Finally, Lachninae were found at a basal position in the mitochondrial tree, whereas they appeared related to Greenideinae in both *Buchnera* and aphid nuclear DNA trees. No a priori can be found in the taxonomic literature on the relatedness of these two subfamilies but interestingly they show similarities in male genitalia morphology (Wieczorek *et al.* 2012).

Below the subfamily level, our analyses retrieved the monophyly of all the tribes represented, except for Macrosiphini, as *Cavariella*, *Pterocomma* genera, and *Muscaphis* are positioned as sister-groups of a clade made of Aphidini and the remaining Macrosiphini. The non-monophyly of Macrosiphini was already suggested by two Aphididae phylogenies, including Macrosiphini representatives (Novakova *et al.* 2013, Hardy *et al.* 2022), and a study dedicated to Macrosiphini (Choi *et al.* 2018). Using different samplings and sets of DNA markers, all three studies show that Macrosiphini are not monophyletic and that *Cavariella* and/or *Pterocomma* (depending on species’ sampling) are allied and form a separate clade, probably with several other Aphidinae genera.

Altogether, our reconstructions validate the classification of aphids into subfamilies as proposed by Remaudière and Remaudière (1997). This alignment with the existing classification is unsurprising, given its recognition as a consensus framework. Concerning tribe delimitations, as exemplified by results in Macrosiphini, a thorough investigation of each subfamily, with dense sampling in each tribe will be needed to reach firm conclusions. Beyond the subfamily level, our results offer limited support for higher level phylogenetic groupings, apart from the clusters highlighted in Table 1. This overall lack of resolution parallels the historical trends within the taxonomic literature, marked by controversies surrounding higher level classifications and aligns well with the idea that aphids have undergone a rapid radiation (von Dohlen and Moran 2000, Heie 2004, Szwedo and Nel 2011).

Conflicts between datasets: is *Buchnera* telling the truth?

Buchnera phylogeny is fully resolved but it conflicts with the aphid topology. There is one striking disagreement between aphid and *Buchnera* topologies: the relative position of Aphidinae. In the *Buchnera* phylogenetic hypothesis, this subfamily appears as a sister-group to other aphid subfamilies, while in aphid dataset analyses, it appears in a more derived position (it clusters with Chaitophorinae/Drepanosiphinae and also Calaphidinae/Phyllaphidinae in most analyses). These results imply that: (i) *Buchnera aphidicola* has not systematically co-specified with its aphid host; or (ii) the aphid phylogenies are too poorly resolved in their deep nodes to be used as solid phylogenetic hypotheses; or (iii) *Buchnera* phylogeny is plagued by artefacts and does not accurately reflect aphid subfamily relationships.

The first hypothesis goes against all evidence accumulated in the literature on *Buchnera* vertical transmission (e.g. Buchner 1965a, Hinde 1971, Wilkinson *et al.* 2003, Koga *et al.* 2012) and previous phylogenetic evidence of aphid/*Buchnera* co-speciation at different evolutionary scales (e.g. Clark *et al.*

2000, Jousselin *et al.* 2009, Xu *et al.* 2018). Moreover, the current study also shows that within subfamilies, *Buchnera* and aphid phylogenetic histories are perfectly congruent (Fig. 5). The lack of co-speciation between aphids and *Buchnera* in deep nodes would imply that horizontal transfers, if any, have only occurred during the early diversification of aphids or that *Buchnera* was acquired independently in different aphid lineages. This latter hypothetical scenario would suppose that aphids had initially diversified fixing different nutritional symbionts (as observed in aphid sister group Adelgidae; von Dohlen *et al.* 2017), recurrently recruiting a *Buchnera*-like endosymbiont (as observed for *Serratia symbiotica* as a co-obligate partner of aphids; Manzano-Marín *et al.* 2023). Current aphid lineages would be the ones that kept *Buchnera*, while others became extinct. While this alternative appears more biologically realistic than horizontal transfers of *Buchnera*, it still seems unlikely considering that almost all aphids host *Buchnera* and that those are almost perfectly syntenic (Chong *et al.* 2019, Manzano-Marín *et al.* 2023). Such synteny is never observed in study systems with multiple symbiont acquisitions (Dial *et al.* 2021) and is a solid marker of a unique acquisition event of *Buchnera*. Given uncertainties in host and symbiont trees, we favour a scenario in which lack of convergence in host and symbiont trees is caused by phylogenetic irresolution. We have, therefore, knowingly chosen not to conduct reconciliation analyses between aphids and *Buchnera* to explore horizontal transfers and multiple acquisition scenarios.

An alternative hypothesis for the lack of parallelism of aphid and *Buchnera* phylogenies is that deep nodes in the aphid trees are unreliable, while *Buchnera* gives a more solid phylogenetic hypothesis. As mentioned previously, the mitochondrial deep nodes are poorly resolved. In addition, our tree is very different from the one obtained by Chen *et al.* (2017), though we employed very similar analytical tools but on an extended taxon sampling. This suggests that phylogenetic resolution using mitogenome data is very sensitive to taxon sampling, which usually indicates difficulties in estimating substitution models (Bernot *et al.* 2023). The nuclear DNA data does not yield a more stable hypothesis. On the other hand, *Buchnera* topology is consistent across all analyses and both Chong *et al.* (2019) and Manzano-Marín *et al.* (2023), conducted on a smaller sampling, yield very similar hypotheses about subfamily relationships. *Buchnera* evolves quickly (Moran *et al.* 1995, Moran 1996) as other primary symbionts of arthropods (Woolfit and Bromham 2003, Degnan *et al.* 2004, Arab *et al.* 2020). If a rapid radiation has indeed given rise to aphid subfamilies, there might have been limited time for informative substitutions to be fixed between speciation events in the aphid nuclear and mitochondrial genomes, while the fast evolutionary rate of *Buchnera* might have allowed their retention. Following this argument, the *Buchnera* phylogeny could represent a more solid phylogenetic hypothesis than aphid phylogenies.

However, the affinity of Aphidinae with Calaphidinae/Phyllaphidinae and Chaitophorinae/Drepanosiphinae is also observed in two recent phylogenomic investigations (Hardy *et al.* 2022, Owen and Miller 2022) (Fig. 1). Though it is hard to be fully confident of the robustness of this result [the support is low in Owen *et al.*'s study and it is not given in Hardy *et al.* (2022)], its consistency across studies suggests that it is not caused by a sampling bias or limited nuclear marker sampling. Recent

investigation on male morphs and male genitalia also validates this clustering (Wieczorek *et al.* 2011): all five subfamilies share normal-sized males without modified genitalia. Furthermore, in the taxonomic literature, Aphidinae are consistently considered as a more derived subfamily (Heie 1985, Shaposhnikov 1981). This stems from its scarcity in the fossil record, as well as biological characteristics such as associations with Rosaceae and diversification on herbaceous hosts. Though these arguments can be discussed, they cast doubt on the hypothesis that *Buchnera* 'is telling the truth', i.e. proposing a solid phylogenetic hypothesis for its hosts. There remains the possibility that some properties of *Buchnera* genomes alter phylogenetic reconstruction efforts. *Buchnera* genomes exhibit strong compositional biases. The use of mixture substitution models is supposed to alleviate the problem of treating such data (Lartillot and Philippe 2004, Lartillot *et al.* 2007, Rodriguez-Ezpeleta *et al.* 2007). However, we have no estimate of their absolute goodness of fit to our data and they can remain inadequate and still too simplistic compared to the evolutionary process driving endosymbiont molecular evolution. In addition, the multiple occurrences of dual-symbioses (Manzano-Marín *et al.* 2023) might further impact *Buchnera* genome evolution. Though our tests for shifts in evolutionary rates between branches of the *Buchnera* tree were not significant, our analyses are preliminary and exclude certain aphid lineages in which symbiotic status is uncertain (Greenideinae and some Hormaphidinae) (Fig. 4). In Hormaphidinae, we have evidence that at least one species has a co-obligate symbiont (Yorimoto *et al.* 2022) and some species have completely lost *Buchnera* (Fukatsu *et al.* 1994, Vogel and Moran 2013). In Greenideinae, the occurrence of a dual-symbiotic system has not been investigated, but the *Buchnera* draft genomes presented here show a highly biased composition that could indicate further *Buchnera* degeneration (Supporting Information, Table S4). More information on the occurrence of dual-symbiotic systems across and within subfamilies is, therefore, needed to conduct thorough tests of evolutionary rate shifts in *Buchnera*. In any case, given the *Buchnera* erratic pattern of genomic erosion, we cannot confidently affirm that it can be relied on to reconstruct robust evolutionary scenarios about aphid subfamily relationships.

CONCLUSION

Heie stated in 1980 that there were 'as many aphid classifications as taxonomists'. Paraphrasing this, we could say that there are 'as many phylogenies as phylogeneticists'. Can we obtain a more solid phylogeny with more data, i.e. more species with a representation of all subfamilies/tribes and genome scale data? We can hope so. However, genome-scale data can also increase the probability of observing incongruent signals (e.g. Cruaud *et al.* 2024). In addition, data errors, miss-assembly, and contaminations often plague phylogenomic analyses (Jeffroy *et al.* 2006, Simion *et al.* 2020). When no a priori hypotheses on phylogenetic relationships are expected, as in aphids, such artefacts are hard to track. Therefore, we want to urge for caution in interpretation of future aphid phylogenomic hypotheses. Their validity in light of morphological characters will need to be evaluated. Interestingly, all strongly sustained subfamily relationships in our study are supported by shared male genitaliae morphology. Therefore, as underlined by Wieczorek *et al.*

(2012), male morphology might actually be a good marker of aphid relatedness. The advent of chromosome-level assemblies could also bring new phylogenetic information, such as chromosomal architectural changes (as in: [Schultz et al. 2023](#)). Finally, if the aphid initial diversification is actually the result of a fast radiation, as for numerous other groups, such as birds to cite one that is not short of phylogenomic investigation ([Suh 2016](#), [Reddy et al. 2017](#)), their early diversification will be difficult to solve.

On a positive side, we consolidate several results: the clustering of Drepanosiphinae with Chaitophorinae, the clustering of Calaphidinae with Phyllaphidinae, the clustering of Hormaphidinae, Thelaxinae, and Anoeciinae ([Table 1](#)). If we follow Remaudière's rationale when building aphid's classification, these three clusters could be worth being ranked as subfamilies. We further validate the paraphyly of Eriosomatinae and their lack of phylogenetic affinity with Hormaphidinae. Given these results, we confirm that two aphid biological characteristics considered as complex host-alternation (present in Anoeciinae, Hormaphidinae, Eriosomatinae, and Aphidinae) and gall-forming (present in Eriosomatinae and Hormaphidinae) have evolved independently in several aphid subfamilies. Finally, we have developed here a useful set of nuclear DNA markers that appear useful for solving relationships within aphid subfamilies. This means that some of the questions that still hover in aphid evolutionary biology—evolution of gall-making, role of host plant shifts, and biogeographic history in species' diversification, and evolution of dual symbioses—can be addressed in a strong phylogenetic framework within some of the highly diversified aphid subfamilies.

SUPPLEMENTARY DATA

Supplementary data are available at *Zoological Journal of the Linnean Society* online.

ACKNOWLEDGEMENTS

We would like to acknowledge the talented artist/scientist Jorge Mariano Collantes Alegre for the aphid cartoons. We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrénées (Bioinfo Genotoul) and the CBGP-HPC computational platform for providing help and/or computing and/or storage resources.

FUNDING

This work was supported by the Marie-Curie AgreenSkills+ fellowship programme co-funded by the EU's *Seventh Framework Programme* (FP7-609398) to A.M.M., the *Marie-Sklodowska-Curie H2020 Programme* (H2020-MSCA-IF-2016, doi: [10.3030/746189](https://doi.org/10.3030/746189)) to E.J., the France Génomique National Infrastructure, funded as part of the Investissements d'avenir programme managed by the Agence Nationale pour la Recherche (ANR-10-INBS-0009) to E.J., C.C., and V.B., and INRAE department SPE project 'Impact Phyto' (2016-2019) to E.J. Data used in this work were (partly) produced through the GenSeq technical facilities of the Institut des Sciences de l'Évolution de Montpellier with the support of LabEx CeMEB, an ANR Investissements d'avenir program (ANR-10-LABX-04-01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONTRIBUTION STATEMENT

E.J.: conceptualization, formal analysis, investigation, writing—original draft, funding acquisition. A.C.D.A.: conceptualization, investigation, writing—review and editing. A.L.C.: investigation, writing—review and editing. M.G.: conceptualization, investigation, writing—review and editing. C.C.: investigation, writing—review and editing. V.B.: investigation, writing—review and editing. A.M.M.: conceptualization, formal analysis, investigation, writing—review and editing, funding acquisition.

CONFLICT OF INTEREST

The authors declare no conflict or competing interests.

DATA AVAILABILITY

The phylogenetic datasets underlying this article are available in Zenodo and can be accessed at <https://doi.org/10.5281/zenodo.10513209>. Available ENA or NCBI accession numbers for mitogenomes and Buchnera genomes are indicated in the Supporting Information, [Table S1](#), they have been submitted within the BioProject PRJEB74807 Genome sequencing of the endosymbiotic consortia and mitochondria of Aphididae (CBGP—INRAE/UniVie). Draft Buchnera genomes or newly assembled mitogenomes from data obtained in other studies are available in archive3 in Zenodo.

REFERENCES

- Arab DA, Bourguignon T, Wang Z *et al.* Evolutionary rates are correlated between cockroach symbionts and mitochondrial genomes. *Biology Letters* 2020;**16**:20190702. <https://doi.org/10.1098/rsbl.2019.0702>
- Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 2012;**19**:455–77. <https://doi.org/10.1089/cmb.2012.0021>
- Bernot JP, Owen CL, Wolfe JM *et al.* Major revisions in pancrustacean phylogeny and evidence of sensitivity to taxon sampling. *Molecular Biology and Evolution* 2023;**40**:msad175. <https://doi.org/10.1093/molbev/msad175>
- Bernt M, Donath A, Jühling F *et al.* MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 2013;**69**:313–9. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Blackman RL, Eastop VF. *Aphids on the World's Crops: An Identification and Information Guide*, 2nd edn. NY: Wiley and sons, 2000, 250.
- Blackman RL, Eastop VF (eds). *Aphids on the World's Herbaceous Plants and Shrubs*. John Wiley & Sons Ltd, **2 vols**. London, UK: The Natural History Museum, 2006, 1439.
- Börner C. *Europae centralis Aphides*. Die blattläuse mitteleuropas. Namen, synonyme, wirtspflanzen, generationszyklen. *Mitteilungen der Thüringischen Botanischen Gesellschaft*. 1952;**3**:1–488.
- Brown JH. Why are there so many species in the tropics?. *Journal of Biogeography* 2014;**41**:8–22. <https://doi.org/10.1111/jbi.12228>
- Buchner P. Symbiosis in animals which suck plant juices. *Endosymbiosis of Animals with Plant Microorganisms*. NY: Interscience, 1965a, 210–432.
- Buchner P. *Endosymbiosis of Animal with Plant Microorganisms*. NY: Interscience, 1965b.
- Cameron SL, Lambkin CL, Barker SC *et al.* A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Systematic Entomology* 2007;**32**:40–59. <https://doi.org/10.1111/j.1365-3113.2006.00355.x>
- Chang H, Qiu Z, Yuan H *et al.* Evolutionary rates of and selective constraints on the mitochondrial genomes of Orthoptera insects

- with different wing types, *Molecular Phylogenetics and Evolution* 2020;**145**:106734. <https://doi.org/10.1016/j.ympev.2020.106734>
- Chen J, Jiang LY, Qiao GX. A total-evidence phylogenetic analysis of Hormaphidinae (Hemiptera: Aphididae), with comments on the evolution of galls. *Cladistics* 2014;**30**:26–66. <https://doi.org/10.1111/cla.12024>
- Chen R, Favret C, Jiang L *et al.* An aphid lineage maintains a bark-feeding niche while switching to and diversifying on conifers. *Cladistics* 2016;**32**:555–72. <https://doi.org/10.1111/cla.12141>
- Chen J, Wang Y, Jiang L *et al.* Mitochondrial genome sequences effectively reveal deep branching events in aphids (Insecta: Hemiptera: Aphididae). *Zoologica scripta* 2017;**46**:706–17. <https://doi.org/10.1111/zsc.12241>
- Choi H, Shin S, Jung S *et al.* Molecular phylogeny of Macrosiphini (Hemiptera: Aphididae): an evolutionary hypothesis for the Pterocomma-group habitat adaptation. *Molecular Phylogenetics and Evolution* 2018;**121**:12–22. <https://doi.org/10.1016/j.ympev.2017.12.021>
- Chong RA, Park H, Moran NA. Genome evolution of the obligate endosymbiont *Buchnera aphidicola*. *Molecular Biology and Evolution* 2019;**36**:1481–9. <https://doi.org/10.1093/molbev/msz082>
- Clark MA, Moran NA, Bauman P *et al.* Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution* 2000;**54**:517–25.
- Coeur d'acier A, Cruaud A, Artige E *et al.* DNA barcoding and the associated PhylAphidB@se website for the identification of European aphids (Insecta: Hemiptera: Aphididae). *PLoS One* 2014;**9**:e97620. <https://doi.org/10.1371/journal.pone.0097620>
- Cruaud A, Rasplus JY, Zhang J *et al.* The Chalcidoidea bush of life: evolutionary history of a massive radiation of minute wasps. *Cladistics* 2024;**40**:34–63. <https://doi.org/10.1111/cla.12561>
- Degnan PH, Lazarus AB, Brock CD *et al.* Host–symbiont stability and fast evolutionary rates in an ant–bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus Blochmannia*. *Systematic Biology* 2004;**53**:95–110. <https://doi.org/10.1080/10635150490264842>
- Delmotte F, Leterme N, Bonhomme J *et al.* Multiple routes to asexuality in an aphid species. *Proceedings Biological Sciences* 2001;**268**:2291–9. <https://doi.org/10.1098/rspb.2001.1778>
- Dial DT, Weglarz KM, Aremu AO *et al.* Transitional genomes and nutritional role reversals identified for dual symbionts of adelgids (Aphidoidea: Adelgidae). *The ISME Journal* 2021;**16**:642–54. <https://doi.org/10.1038/s41396-021-01102-w>
- Dixon AFG, Kindlmann P, Leps J *et al.* Why there are so few species of aphids, especially in the tropics. *The American Naturalist* 1987;**129**:580–92. <https://doi.org/10.1086/284659>
- von Dohlen CD. Aphid molecular systematics: history, progress and prospects. *Redia* 2009;**92**:34–45.
- von Dohlen CD, Moran NA. Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biological Journal of the Linnean Society* 2000;**71**:689–717.
- von Dohlen CD, Spaulding U, Patch KB *et al.* Dynamic acquisition and loss of dual-obligate symbionts in the plant-sap-feeding adelgidae (Hemiptera: Sternorrhyncha: Aphidoidea). *Frontiers in Microbiology* 2017;**8**:1037. <https://doi.org/10.3389/fmicb.2017.01037>
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 2019;**20**:1–14.
- Fan L, Wu D, Goremykin V *et al.* Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nature Ecology & Evolution* 2020;**4**:1213–9. <https://doi.org/10.1038/s41559-020-1239-x>
- Favret C 2024 *Aphid Species File*, v.5.0/5.0 (12 December 2023, date last accessed). <http://Aphid.SpeciesFile.org>
- Feng H, Edwards N, Anderson CMH *et al.* Trading amino acids at the aphid–*Buchnera* symbiotic interface. *Proceedings of the National Academy of Sciences* 2019;**116**:16003–11. <https://doi.org/10.1073/pnas.1906223116>
- Fukatsu T, Aoki S, Kurosu U *et al.* Phylogeny of Cerataphidini aphids revealed by their symbiotic microorganisms and basic structure of their galls: implications for host–symbiont coevolution and evolution of sterile soldier castes. *Zoological Science* 1994;**11**:612–23.
- Galan M, Pons JB, Tournayre O *et al.* Metabarcoding for the parallel identification of several hundred predators and their prey: application to bat species diet analysis. *Molecular Ecology Resources* 2018;**18**:474–89. <https://doi.org/10.1111/1755-0998.12749>
- Ghosh AK. Hormaphidinae: distribution, phylogeny and systematics. In: *Evolution and Biosystematics of Aphids. Proceedings of the International Aphidological Symposium at Jablonna*, 5–11 April, 1981. Warsaw, Poland: Institute of Zoology, Polish Academy of Sciences, 1985, 303–36.
- Gil R, Sabater-Munoz B, Latorre A *et al.* Extreme genome reduction in *Buchnera* spp.: Toward the minimal genome needed for symbiotic life. *Proceedings of the National Academy of Sciences of the United States of America* 2002;**99**:4454–8. <https://doi.org/10.1073/pnas.062067299>
- Hardy NB, Kaczvinsky C, Bird G *et al.* Geographic isolation drives speciation in Nearctic aphids. *Communications Biology* 2022;**5**:796. <https://doi.org/10.1038/s42003-022-03771-5>
- Heie OE. Studies on fossil aphids (Homoptera: Aphidoidea), *Spolia Zoologica Musei Hauniensis*, 1967;**26**:1–273.
- Heie OE. *The Aphidoidea of Fennoscandia and Denmark I. Mindaridae, Hormaphididae, Thelaxidae, Anoecidae and Pemphigidae.* (*Fauna entomologica scand.*, 9) (1980). Leiden, The Netherlands: Brill, 1980, 236.
- Heie OE. Fossil Aphids. In: *Evolution and Biosystematics of Aphids. Proceedings of the International Aphidological Symposium at Jablonna*, 5–11 April, 1981. Warsaw, Poland: Institute of Zoology, Polish Academy of Sciences, 1985, 101–34.
- Heie OE. 'Paleontology and phylogeny'. In: Minks AK, Harrewijn P (eds), *Aphids. Their Biology, Natural Enemies and Control*, Vol. **2a**. Amsterdam: Elsevier, 1987, 367–91.
- Heie OE. Why are there so few aphid species in the temperate areas of the southern-hemisphere. *European Journal of Entomology* 1994;**91**:127–33.
- Heie OE. The history of the studies on aphid palaeontology and their bearing on the evolutionary history of aphids. In: Simon J-C *et al.* (eds), *Aphids in a New Millennium*. Paris: INRA Editions, 2004, 151–58.
- Hinde R. The control of the mycetome symbiotes of the aphids *Brevicoryne brassicae*, *Myzus persicae*, and *Macrosiphum rosae*. *Journal of Insect Physiology* 1971;**17**:1791–800. [https://doi.org/10.1016/0022-1910\(71\)90076-x](https://doi.org/10.1016/0022-1910(71)90076-x)
- Husnik F, Chrudimsky T, Hypsa V. Multiple origins of endosymbiosis within the Enterobacteriaceae (gamma-Proteobacteria): convergence of complex phylogenetic approaches. *BMC Biology* 2011;**9**:9.
- Hyatt D, Chen GL, LoCascio PF *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;**11**:119. <https://doi.org/10.1186/1471-2105-11-119>
- Ilharco F, Harten V. Systematic. In: Minks AK, Harrewijn P (eds), *Aphids, their Biology, Natural Enemies and Control*, Vol. **2a**. Amsterdam: Elsevier, 1987, 51–77.
- Jaquiere J, Stoeckel S, Larose C *et al.* Genetic control of contagious asexuality in the pea aphid. *PLoS Genetics* 2014;**10**:e1004838. <https://doi.org/10.1371/journal.pgen.1004838>
- Jeffroy O, Brinkmann H, Delsuc F *et al.* Phylogenomics: the beginning of incongruence? *Trends in Genetics* 2006;**22**:225–31. <https://doi.org/10.1016/j.tig.2006.02.003>
- Jousselin E, Elias M. Testing host-plant driven speciation in phytophagous insects: a phylogenetic perspective. 2019:arXiv. 1910.09510, ver. 1 [peer-reviewed and recommended by *PCI Evolutionary Biology*]. <https://arxiv.org/abs/1910.09510v1>
- Jousselin E, Desdésives Y, Coeur d'acier A. Fine-scale cospeciation between *Brachycaudus* and *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in aphids. *Proceedings Biological Sciences* 2009;**276**:187–96. <https://doi.org/10.1098/rspb.2008.0679>

- Jousselin E, Genson G, Coeur d'Acier A. Evolutionary lability of a complex life cycle in the aphid genus *Brachycaudus*. *BMC Evolutionary Biology* 2010;10:10.
- Kalyaanamoorthy S, Minh BQ, Wong TKF *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017;14:587–9. <https://doi.org/10.1038/nmeth.4285>
- Katoh K, Standley DM. MAFFT multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>
- Koga R, Meng X-Y, Tsuchida T *et al.* Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte–embryo interface. *Proceedings of the National Academy of Sciences* 2012;109:E1230–E37.
- Kowalczyk A, Meyer WK, Partha R *et al.* RERconverge: an R package for associating evolutionary rates with convergent traits. *Bioinformatics* 2019;35:4815–7. <https://doi.org/10.1093/bioinformatics/btz468>
- Lamelas A, Gosalbes MJ, Moya A *et al.* New clues about the evolutionary history of metabolic losses in bacterial endosymbionts, provided by the genome of *Buchnera aphidicola* from the Aphid *Cinara tujafilina*. *Applied and Environmental Microbiology* 2011;77:4446–54. <https://doi.org/10.1128/AEM.00141-11>
- Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* 2004;21:1095–109. <https://doi.org/10.1093/molbev/msh112>
- Lartillot N, Brinkmann H, Philippe H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evolutionary Biology* 2007;7:S4. <https://doi.org/10.1186/1471-2148-7-s1-s4>
- Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 2009;25:2286–8. <https://doi.org/10.1093/bioinformatics/btp368>
- Lee Y, Kanturski M, Foottit RG *et al.* Molecular phylogeny and evolution of Calaphidinae (Hemiptera: Aphididae). *Cladistics* 2022;38:159–86. <https://doi.org/10.1111/cla.12487>
- Lukasiak P, Chong RA, Nazario K *et al.* One hundred mitochondrial genomes of cicadas. 2019;110:247–56. <https://doi.org/10.1093/jhered/esy068>
- Mackauer M. Parasitological data as an aid in aphid classification. *Canadian Entomologist* 1965;97:1016–24. <https://doi.org/10.4039/ent971016-10>
- Mackenzie A, Dixon AFG. Host alternation in aphids: constraint versus optimization. *The American Naturalist* 1990;136:132–4. <https://doi.org/10.1086/285086>
- Manzano-Marín A, Coeur d'acier A, Clamens A-L *et al.* A freeloader? The highly eroded yet large genome of the *Serratia symbiotica* symbiont of *Cinara strobilifera*. *Genome Biology and Evolution* 2018;10:2178–89. <https://doi.org/10.1093/gbe/evy173>
- Manzano-Marín A, Coeur d'acier A, Clamens A-L *et al.* Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *The ISME Journal* 2020;14:259–73. <https://doi.org/10.1038/s41396-019-0533-6>
- Manzano-Marín A, Coeur d'acier A, Clamens A-L *et al.* Co-obligate symbioses have repeatedly evolved across aphids, but partner identity and nutritional contributions vary across lineages. *Peer Community Journal* 2023;3:e46. <https://doi.org/10.24072/pcjournal.278>
- Martin J. Creating error-proof indexes for high throughput sequencing. *Zenodo* 2019. <https://doi.org/10.5281/zenodo.3350207>
- Meseguier AS, Manzano-Marín A, Coeur d'acier A *et al.* *Buchnera* has changed flatmate but the repeated replacement of co-obligate symbionts is not associated with the ecological expansions of their aphid hosts. *Molecular Ecology* 2017;26:2363–78. <https://doi.org/10.1111/mec.13910>
- Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* 2020;37:2727–33. <https://doi.org/10.1093/molbev/msaa106>
- Misof B, Liu S, Meusemann K *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* 2014;346:763–7. <https://doi.org/10.1126/science.1257570>
- Miyazaki M. Forms and morphs of aphids. In: Minks AK, Harrewijn P (eds), *Aphids, their Biology, Natural Enemies and Control*. Amsterdam: Elsevier, 1987, 163–95.
- Monnin D, Jackson R, Kiers ET *et al.* Parallel evolution in the integration of a co-obligate aphid symbiosis. *Current Biology* 1949;30:1949–57. e6. <https://doi.org/10.1016/j.cub.2020.03.011>
- Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 1996;93:2873–8. <https://doi.org/10.1073/pnas.93.7.2873>
- Moran N, Mira A. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biology* 2001;2:1–12.
- Moran NA, vonDohlen CD, Baumann P. Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. *Journal of Molecular Evolution* 1995;41:727–31.
- Mullen KM, Schneider SA, Normark BB. New single-copy nuclear genes for use in scale insect systematics. *Redia* 2017;99:207.
- Munson MA, Baumann P, Clark MA *et al.* Evidence for the establishment of aphid–eubacterium endosymbiosis in an ancestor of four aphid families. *Journal of Bacteriology* 1991;173:6321–4. <https://doi.org/10.1128/jb.173.20.6321-6324.1991>
- Nguyen L-T, Schmidt HA, von Haeseler A *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 2014;32:268–74. <https://doi.org/10.1093/molbev/msu300>
- Novakova E, Hypsa V, Klein J *et al.* Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Molecular Phylogenetics and Evolution* 2013;68:42–54. <https://doi.org/10.1016/j.ympev.2013.03.016>
- Ortiz-Rivas B, Moya A, Martínez-Torres D. Molecular systematics of aphids (Homoptera: Aphididae): new insights from the long-wavelength opsin gene. *Molecular Phylogenetics and Evolution* 2004;30:24–37. [https://doi.org/10.1016/s1055-7903\(03\)00155-6](https://doi.org/10.1016/s1055-7903(03)00155-6)
- Ortiz-Rivas B, Martínez-Torres D. 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* 2010;55:305–17. <https://doi.org/10.1016/j.ympev.2009.12.005>
- Owen CL, Miller GL. Phylogenomics of the Aphididae: deep relationships between subfamilies clouded by gene tree discordance, introgression and the gene tree anomaly zone. *Systematic Entomology* 2022;47:470–86. <https://doi.org/10.1111/syen.12542>
- Podsiadlowski L, Vilcinskis A. Phylogeny of the aphids. In: Vilcinskis A. (ed.), *Biology and Ecology of Aphids*. Boca Raton, Florida, USA: CRC Press, 2016, 1–13.
- Powell G, Hardie J. The chemical ecology of aphid host alternation: how do return migrants find the primary host plant? *Applied Entomology and Zoology* 2001;36:259–67. <https://doi.org/10.1303/aez.2001.259>
- Qiao GX, Zhang HC. Molecular phylogeny of Pemphiginae (Hemiptera: Aphididae) inferred from nuclear gene EF-1 α sequences. *Bulletin of Entomological Research* 2008;98:499–507.
- Quednau FW. *Atlas of the Drepanosiphine Aphids of the World. Part III: Mindarinae Tullgren 1909 to Saltusaphidinae Baker 1920 (Hemiptera: Sternorrhyncha, Aphididae)*. Gainesville, Florida, USA: Memoirs of the American Entomological Institute, 2010, 361.
- Reddy S, Kimball RT, Pandey A *et al.* Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Systematic Biology* 2017;66:857–79. <https://doi.org/10.1093/sysbio/syx041>
- Remaudière G, Remaudière M. *Catalogue des Aphididae du Monde—Catalogue of the World's Aphididae (Homoptera, Aphidoidea)*. INRA. Institut National de la Recherche Agronomique. Paris: Techniques et Pratiques, 1997, 376.
- Remaudière G, Stroyan H. Un *Tamalia* Nouveau de Californie (USA) Discussion sur les Subfam. Nov. (Hom. Aphididae). *Les Annales de la Société entomologique de France (N.S.)* 1984;20:93–103.

- Renoz F, Ambroise J, Bearzatto B *et al.* The di-symbiotic systems in the Aphids *Sipha maydis* and *Periphyllus lyropictus* provide a contrasting picture of recent co-obligate nutritional endosymbiosis in Aphids. *Microorganisms* 2022;**10**:1360. <https://doi.org/10.3390/microorganisms10071360>
- Rodriguez-Ezpeleta N, Brinkmann H, Roure B *et al.* Detecting and overcoming systematic errors in genome-scale phylogenies. *Systematic Biology* 2007;**56**:389–99. <https://doi.org/10.1080/10635150701397643>
- Russell LM. The genus *Neophyllaphis* and its species (Hemiptera: Homoptera: Aphididae). *Florida Entomologist* 1982;**65**:538–73. <https://doi.org/10.2307/3494690>
- Sabater-Muñoz B, Toft C, Alvarez-Ponce D *et al.* Chance and necessity in the genome evolution of endosymbiotic bacteria of insects. *The ISME Journal* 2019;**11**:1291–304. <https://doi.org/10.1038/ismej.2017.18>
- Schultz DT, Haddock SHD, Bredeson JV *et al.* Ancient gene linkages support ctenophores as sister to other animals. *Nature* 2023;**618**:110–7. <https://doi.org/10.1038/s41586-023-05936-6>
- Shaposhnikov G. Host specificity and adaptation to new hosts in aphids (Homoptera, Aphidoidea) in the process of natural selection (Experimental research). *Entomological Review* 1961;**40**:412–26.
- Shaposhnikov GC. The main features of the evolution of aphids. *International Aphidological Symposium at Jablonna*. Jablonna: Polska Akademia Nauk, Warsaw, 1981, 19–99.
- Shaposhnikov GC. Evolution of aphids in relation to evolution of plants. In: Minks AK, Harrewin P (eds), *Aphids. Their Biology, Natural Enemies and Control*. Amsterdam: Elsevier, 1987, 409–13.
- Shimodaira H. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 2002;**51**:492–508. <https://doi.org/10.1080/10635150290069913>
- Simion P, Delsuc F, Philippe H. *To What Extent Current Limits of Phylogenomics Can Be Overcom. Phylogenetics in the Genomic Era 2020:2–1* [no commercial publisher] Authors open access book].
- Smith TE, Moran NA. Coordination of host and symbiont gene expression reveals a metabolic tug-of-war between aphids and *Buchnera*. *Proceedings of the National Academy of Sciences of the United States of America* 2020;**117**:2113–21. <https://doi.org/10.1073/pnas.1916748117>
- Sow A, Brévault T, Benoit L *et al.* Deciphering host–parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding. *Scientific Reports* 2019;**9**:3646. <https://doi.org/10.1038/s41598-019-40243-z>
- Suh A. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zoologica Scripta* 2016;**45**:50–62. <https://doi.org/10.1111/zsc.12213>
- Szwedo J, Nel A. The oldest aphid insect from the Middle Triassic of the Vosges, France. *Acta Palaeontologica Polonica* 2011;**56**:757–66. <https://doi.org/10.4202/app.2010.0034>
- Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 2007;**56**:564–77. <https://doi.org/10.1080/10635150701472164>
- Toft C, Fares MA. Selection for translational robustness in *Buchnera aphidicola*, endosymbiotic Bacteria of Aphids. *Molecular Biology and Evolution* 2009;**26**:743–51. <https://doi.org/10.1093/molbev/msn301>
- Vogel KJ, Moran NA. Functional and evolutionary analysis of the genome of an obligate fungal symbiont. *Genome Biology and Evolution* 2013;**5**:891–904. <https://doi.org/10.1093/gbe/evt054>
- von Dohlen CD. Macroevolution of aphid-plant relationships: an integrated approach. In: Simon JC, Dedryver CA, Rispe C, Hüllé M, editors. *Aphids in a New Millennium*. Paris: INRA, 2004, 385–90.
- Wahlberg N, Wheat CW. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology* 2008;**57**:231–42. <https://doi.org/10.1080/10635150802033006>
- Wahlberg N, Peña C, Ahola M *et al.* PCR primers for 30 novel gene regions in the nuclear genomes of Lepidoptera. *ZooKeys* 2016;**596**:129–41. <https://doi.org/10.3897/zookeys.596.8399>
- Wang Y, Huang XL, Qiao GX. Comparative analysis of mitochondrial genomes of five aphid species (Hemiptera: Aphididae) and phylogenetic implications. *PLoS One* 2013;**8**:e77511. <https://doi.org/10.1371/journal.pone.0077511>
- Wang H-C, Susko E, Roger AJ. The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference. *Systematic Biology* 2019;**68**:1003–19. <https://doi.org/10.1093/sysbio/syz021>
- Wernegreen JJ, Richardson AO, Moran NA. Parallel acceleration of evolutionary rates in symbiont genes underlying host nutrition. *Molecular Phylogenetics and Evolution* 2001;**19**:479–85. <https://doi.org/10.1006/mpev.2001.0929>
- Wieczorek K, Wojciechowski W. Approaches to classification of Drepanosiphidae aphids (Hemiptera, Aphidoidea: Drepanosiphidae). *Aphids Other Hemipterous Insects* 2007;**13**:7–13.
- Wieczorek K, Plachno BJ, Świątek P. Comparative morphology of the male genitalia of Aphididae (Insecta, Hemiptera): Part 1. *Zoomorphology* 2011;**130**:289–303. <https://doi.org/10.1007/s00435-011-0134-z>
- Wieczorek K, Plachno BJ, Świątek P. A comparative morphology of the male genitalia of Aphididae (Insecta, Hemiptera): part 2. *Zoomorphology* 2012;**131**:303–24. <https://doi.org/10.1007/s00435-012-0163-2>
- Wieczorek K, Lachowska-Cierlik D, Kajtoch L *et al.* The relationships within the Chaitophorinae and Drepanosiphinae (Hemiptera, Aphididae) inferred from molecular-based phylogeny and comprehensive morphological data. *PLoS One* 2017;**12**:e0173608. <https://doi.org/10.1371/journal.pone.0173608>
- Wiegmann BM, Trautwein MD, Kim J-W *et al.* Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biology* 2009;**7**:34. <https://doi.org/10.1186/1741-7007-7-34>
- Wilkinson TL, Fukatsu T, Ishikawa H. Transmission of symbiotic bacteria *Buchnera* to parthenogenetic embryos in the aphid *Acyrtosiphon pisum* (Hemiptera: Aphidoidea). *Arthropod Structure & Development* 2003;**32**:241–5. [https://doi.org/10.1016/S1467-8039\(03\)00036-7](https://doi.org/10.1016/S1467-8039(03)00036-7)
- Wilson ACC, Ashton PD, Calevro F *et al.* Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Molecular Biology* 2010;**19**:249–58. <https://doi.org/10.1111/j.1365-2583.2009.00942.x>
- Wojciechowski W. *Studies on the Systematic System of Aphids (Homoptera, Aphidinea)*. Katowice, Poland: Uniwersytet Śląski, 1992.
- Wojciechowski W, Wieczorek K. The systematic position of Chaitophorinae (Hemiptera, Aphidoidea) in the light of anatomy research. *Insect Systematics & Evolution* 2004;**35**:317–27. <https://doi.org/10.1163/187631204788920158>
- Woolfit M, Bromham L. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Molecular Biology and Evolution* 2003;**20**:1545–55. <https://doi.org/10.1093/molbev/msg167>
- Xu T-T, Chen J, Jiang L-Y *et al.* Historical and cospeciating associations between Cerataphidini aphids (Hemiptera: Aphididae: Hormaphidinae) and their primary endosymbiont *Buchnera aphidicola*. *Zoological Journal of the Linnean Society* 2018;**183**:236–236. <https://doi.org/10.1093/zoolinnean/zly006>
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 2007;**24**:1586–91. <https://doi.org/10.1093/molbev/msm088>
- Yorimoto S, Hattori M, Kondo M *et al.* Complex host/symbiont integration of a multi-partner symbiotic system in the eusocial aphid *Ceratovacuna japonica*. *Iscience* 2022;**25**:105478. <https://doi.org/10.1016/j.isci.2022.105478>
- Zhang H, Liu Q, Lu C *et al.* The first complete mitochondrial genome of Lachninae species and comparative genomics provide new insights into the evolution of gene rearrangement and the repeat region. *Insects* 2021;**12**:55. <https://doi.org/10.3390/insects12010055>