Molecular Phylogenetics and Evolution 55 (2010) 305–317

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

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

Benjamín Ortiz-Rivas, David Martínez-Torres *

Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado de Correos 22085, 46071 Valencia, Spain

ARTICLE INFO

Article history:
Received 31 July 2009
Revised 16 November 2009
Accepted 3 December 2009
Available online 21 December 2009

Keywords:
Aphid
Phylogeny
Molecular data
Combined analysis
Host alternation

ABSTRACT

The first molecular studies on the phylogeny of aphids (Hemiptera: Aphididae) bumped into a striking lack of phylogenetic structure for taxa levels higher than tribe, probably as a consequence of the rapid adaptive radiation that this group of insects went through during the Late Cretaceous. Here we present a new attempt to infer the relationships between major aphid taxa by the separate and combined analysis of two nuclear sequences (the long-wavelength opsin gene and the elongation factor 1α gene) and two mitochondrial sequences (the genes encoding the subunit 6 of the F-ATPase and the subunit II of the cytochrome oxidase). Our results confirm previous results with the grouping of the subfamilies analysed in three main lineages, that are named A + D (subfamilies Aphidinae, Calaphidinae, Chaitophorinae, Drepanosiphinae and Pterocommatinae), E + T (subfamilies Anoeciinae, Eriosomatinae, Hormaphidinae, Mindarinae and Thelaxinae) and L (subfamily Lachninae). Furthermore, phylogenetic reconstructions generally support the early branching of the subfamily Lachninae in the phylogeny of aphids. Although some relationships among subfamilies inside lineages are not highly supported, our results are compatible with a scenario for the evolution of aphid life cycles with only four transitions of feeding from gymnosperms to angiosperms and two origins of host alternation.

1. Introduction

Aphids are a cosmopolitan group of phloem feeding hemipterans comprising more than 4700 species, with higher diversity in the temperate regions of the world (Remaudière and Remaudière, 1997; Nieto-Nafría and Mier-Durante, 1998). The striking complexity and variability of their life cycles, which include cyclical parthenogenesis and sometimes alternation between distantly related plant hosts, and the long lasting symbiotic association with the endobacterium *Buchnera aphidicola*, represent salient aspects of the interesting biology of these insects. The systematics and phylogenetics of this group have been controversial since the first studies on the issue (for a review see Wojciechowski, 1992 or Ilharco and van Harten, 1987). Three families have been generally recognised: Aphididae (comprising the so-called true aphids), with viviparous parthenogenetic females, and Adelgidae and Phylloxeridae, with oviparous parthenogenetic females. Nevertheless, the phylogenetic relationships within the highly diversified Aphididae are not resolved. A phylogeny based on morphological characters was proposed by Heie (1987) (Fig. 1a) and has been extensively referenced thereafter. Several phylogenetic reconstructions inferred from *Buchnera* DNA sequences were compatible with Heie's topology (Munson et al., 1991; Moran et al., 1993; Moran and Baumann, 1994; Rouhbakhsh et al., 1996; Silva et al., 1998; Brynnel et al., 1998; Baumann et al., 1999; van Ham et al., 1999, 2000), validating it and supporting a parallel coevolution of aphids and *Buchnera*. However, many of these studies were far from comprehensive, and strongly biased towards representatives of the subfamily Aphidinae. Different interpretations on the homology of some morphological characters led Wojciechowski to propose an alternative phylogenetic hypothesis (Wojciechowski, 1992) (Fig. 1b).

Attempts to infer the global phylogeny of the family Aphididae with molecular data were firstly made in two reports using mitochondrial DNA sequences (von Dohlen and Moran, 2000; Martínez-Torres et al., 2001). Both studies showed little phylogenetic structure at levels higher than tribe except for the monophyly of the subfamilies Aphidinae and Lachninae. A rapid adaptive radiation at the tribal level connected to a shift from gymnosperms to angiosperms was invoked to interpret the lack of support to the relationships in the deepest nodes of the tree. Although a rapid radiation in the Aphididae presents challenges for determining

* Corresponding author. Fax: +34 963 543 670.
E-mail addresses: benjamin.ortiz@uv.es (B. Ortiz-Rivas), david.martinez@uv.es (D. Martínez-Torres).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.12.005

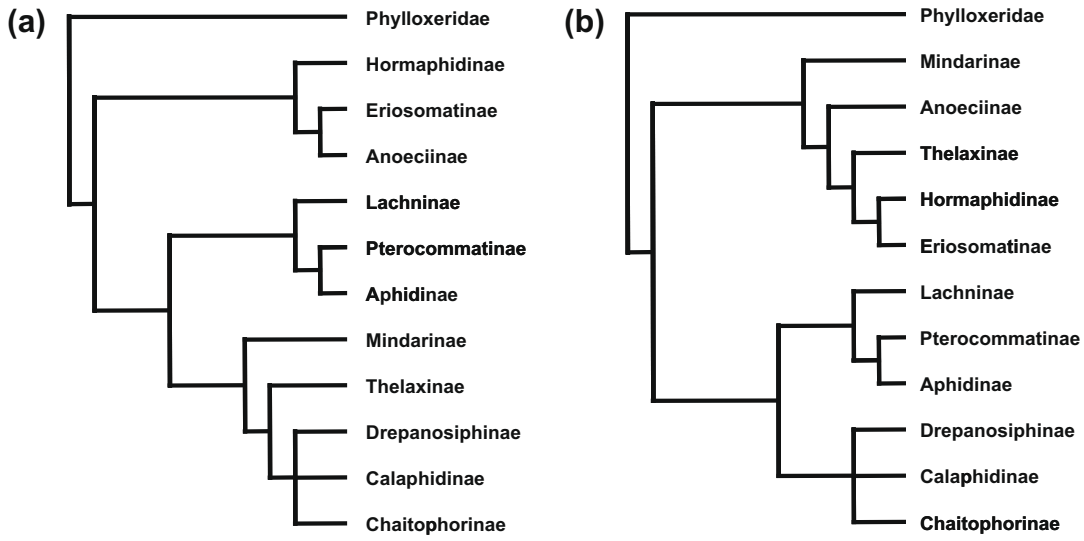

Fig. 1. Morphological phylogenetic hypotheses proposed by (a) Heie (1987) and (b) Wojciechowski (1992). Only those subfamilies included in the present report are shown. Taxa ranks and splits have been adapted to Remaudière and Remaudière (1997).

phylogenetic relationships among its subfamilies, the subsequent use of the nuclear gene encoding the long-wavelength opsin (LWO) (Ortiz-Rivas et al., 2004) allowed some advances in the knowledge of the deepest relationships in the family. The combined use of the LWO gene and the previously sequenced gene encoding the subunit 6 of the mitochondrial F-ATPase (ATP6) revealed the existence of three main aphid lineages, which were named: A + D (including the subfamilies Aphidinae, Calaphidinae, Chaitophorinae, Drepanosiphinae and Pterocommatinae); P + T (containing the subfamilies Anoeciinae, Eriosomatinae, Hormaphidinae and Thelaxinae; hereafter renamed E + T) and L (including the representatives of the subfamily Lachninae). Furthermore, the data analysed suggested a basal position of the subfamily Lachninae in the phylogeny of the family Aphididae. These relationships highly disagreed with Heie's phylogeny of aphids but were more congruent with Wojciechowski's proposal.

Any further improvement in the knowledge of the relationships among major aphid taxa would be of special relevance in several aspects of their biology, like the proposed strict parallel evolution of these insects and the endosymbiont *Buchnera aphidicola* or the evolution of their complex life cycles, including the evolution of feeding and the number of independent origins of host alternation. The obtention of a robust phylogeny of aphids would also be helpful for future genomic research following the recent sequencing of the genome of the pea aphid *Acyrthosiphon pisum* (The International Aphid Genomics Consortium, 2009). The nuclear gene for the translation elongation factor 1α (EF1α) and the mitochondrial gene for the cytochrome oxidase II (COII) have been used in several studies for solving diverse issues in aphid biology and systematics (Stern et al., 1997; Moran et al., 1999; Normark, 2000; von Dohlen et al., 2002, 2006; Inbar et al., 2004). In this report, we present the results of a molecular phylogenetic analysis of a broader survey of aphid species and subfamilies and the compilation and combination of sequences of the nuclear LWO and EF1α genes and the mitochondrial ATP6 and COII genes.

2. Materials and methods

2.1. Aphid species analysed

Aphid species used in this study are listed in Table 1, with indication of the genes analysed for each of them. The classification used in the *Catalogue of the World's Aphididae* (Remaudière and Stroyan, 1984; Remaudière and Quednau, 1988; Quednau and Remaudière, 1994; Remaudière and Remaudière, 1997), with slight modifications by Nieto-Nafría et al. (1997), was followed. Representatives of 11 aphid subfamilies were included: Anoeciinae, Aphidinae, Calaphidinae, Chaitophorinae, Drepanosiphinae, Eriosomatinae, Hormaphidinae, Lachninae, Mindarinae, Pterocommatinae and Thelaxinae. The species *Daktulosphaira vitifoliae*, belonging to the family Phylloxeridae, was used as outgroup.

2.2. DNA extraction and PCR amplification

Total DNA was extracted from single aphids following the method previously described (Martínez et al., 1992) but omitting all the alkali-treatment-related steps. The PCR amplification and sequencing of LWO and ATP6 sequences for the present report were carried out as previously described (Martínez-Torres et al., 2001; Ortiz-Rivas et al., 2004). For the amplification of EF1α sequences, primers were designed from the alignment of available sequences from aphid species belonging to the subfamilies Aphidinae, Lachninae and Hormaphidinae, obtained in previous studies (see Table 1). Most of the sequences were amplified in a single round of PCR, using primers efs175 (Moran et al., 1999) and efr1 (5′GTGTGGCAATSCAANACNGGAGT3′). The reaction was carried out on a GeneAmp PCR System 9700 (Applied Biosystems), with the following conditions: 94 °C for 3 min; 35 cycles of 94 °C for 30 s, 50 °C for 1 min and 68 °C for 1.5 min; and a final extension step of 7 min at 68 °C. In some instances, a nested PCR was necessary, which was accomplished using primers efs1 (5′TGGA CAAAYTKAAGGCTGAACG3′) and efr3 (5′GTRTASCCRTTGGAAATTT GACC3′), and the same conditions used for the first PCR except for increasing the annealing temperature to 52 °C. The reactions were done on approximately 100 ng of DNA and using 0.25 μl of *Taq* polymerase (Eppendorf).

2.3. Sequencing of PCR products

PCR products were purified by ammonium precipitation and reconstituted in 10 μl of standard TE buffer (10 mM Tris, 1 mM EDTA). Sequencing and occasional cloning of LWO and ATP6 genes were done as previously described (Martínez-Torres et al., 2001; Ortiz-Rivas et al., 2004). All EF1α sequences were obtained by di-

Table 1
Aphid species analysed, classified following Remaudière and Remaudière (1997) and EBI Accession Numbers[a].

Subfamily	Tribe	Species	LWO	EF1α	ATP6	COII
Anoeciinae		*Anoecia* sp.	AJ53946319	FM174706	AJ547789[19]	–
Aphidinae	Aphidini	*Aphis fabae*	–	AY219724[7]	–	AM085434[10]
		Aphis nerii	–	FM174700	FM174676	AM085436[10]
		Aphis spiraecola	FM177115	–	–	AM085435[10]
		Melanaphis luzuella	–	–	–	AM085392[10]
		Rhopalosiphum padi	FM177114	FM174699	AJ298673[4]	U36749[16]
		Schizaphis graminum	–	AY219720[7]	–	AF059698[2]
	Macrosiphini	*Acyrthosiphon pisum*	AJ489281[19]	FM174698	AJ298675[4]	AF059699[2]
		Hyperomyzus lactucae	–	–	–	AM085391[10]
		Macrosiphoniella ludovicianae	–	–	–	AF059696[2]
		Megoura viciae	AF189714[12]	FM174696	FM174677	–
		Myzus persicae	AJ489282[19]	FM174697	AJ298674[4]	AF143582[18]
		Sitobion avenae	–	DQ005155[11]	–	U41116[17]
		Uroleucon ambrosiae	–	–	–	AF196376[15]
Calaphidinae	Panaphidini	*Chromaphis juglandicola*	FM177105	FM174691	FM174678	–
		Hoplocallis pictus	AJ539466[19]	FM174693	AJ298668[4]	–
		Panaphis juglandis	FM177106	FM174692	AJ298669[4]	–
		Tinocallis takachihoensis	AJ489286[19]	FM174694	AJ547792[19]	–
Chaitophorinae	Chaitophorini	*Chaitophorus leucomelas*	FM177104	FM174705	AJ298670[4]	AF444288[6]
		Peryphillus bulgaricus	AJ489285[19]	FM174704	AJ547791[19]	AF444290[6,a]
Drepanosiphinae		*Drepanosiphum oregonensis*	AJ489284[19]	FM174695	AJ298671[4]	–
Eriosomatinae	Eriosomatini	*Eriosoma lanuginosum*	AJ539464[19]	FM174709	AJ298665[4]	–
		Tetraneura caerulescens	AJ489291[19]	FM174689	AJ298666[4]	–
	Fordini	*Aploneura lentisci*	AJ489289[19]	FM163601	AJ298663[4]	AY227092[9]
		Baizongia pistaciae	AJ489290[19]	FM163599	AJ547790[19]	AY227093[9]
		Forda marginata	FM177108	FM163596	AM996893	AY227098[9]
		Geoica utricularia	FM177110	FM163600	AJ298662[4]	AY227094[9,b]
		Melaphis rhois	–	–	–	U36747[5]
		Nurudea shiraii	–	–	–	AF454627[5]
		Paracletus cimiciformis	FM177109	FM163597	AM996884	AY227102[9]
		Schlechtendalia chinensis	–	–	–	AF454628[5]
		Slavum wertheimae	–	–	–	AY227103[9]
		Smynthurodes betae	FM177111	FM163598	AM996898	AY227104[9]
	Pemphigini	*Pemphigus populi*	AJ489288[19]	FM163603	AJ547793[19]	L27337[1,c]
		Thecabius sp.	FM177112	FM163602	AM996899	AY182307[8,d]
Hormaphidinae	Cerataphidini	*Cerataphis* sp.	AJ539465[19]	FM174690	FM174679	L27327[14,e]
		Ceratoglyphina bambusae	–	–	–	L27327[14]
		Ceratovacuna japonica	–	–	–	L27328[14]
		Glyphinaphis bambusae	–	–	–	L27331[1]
		Pseudoregma alexanderi	–	–	–	L27335[14]
	Hormaphidini	*Hamamelistes betulinus*	–	AF454597[5]	–	AF328782[5]
		Hormaphis betulae	–	AF454611[5]	–	AF454623[5]
	Nipponaphidini	*Metanipponaphis rotunda*	–	–	–	AF454624[5]
		Neothoracaphis yanonis	–	–	–	L27334[14]
		Nipponaphis distyliicola	–	AF454614[5]	–	L27333[5]
		Sinonipponaphis monzeni	–	AF454615[5]	–	–
Lachninae	Eulachnini	*Cinara cedri*		FM174683		
		Cinara tujafilina	AJ489294[19]	FM174684	AJ298680[4,f]	AF156196[3,g]
		Essigella fusca	–	AF163866[3]	–	AF156190[3]
		Eulachnus rileyi	AJ489292[19]	FM174708	AJ298681[4,h]	AF156191[3]
		Schizolachnus pineti	–	–	AJ298678[4]	–
	Lachnini	*Lachnus roboris*	AJ489293[19]	FM174686	AJ298682[4]	AF156198[3,i]
		Maculolachnus submacula	FM177103	FM174688	AJ298677[4]	AF156200[3]
		Nippolachnus piri	–	AF147811[3]	–	AF156204[3]
		Pterochloroides persicae	FM177102	FM174687	FM174680	–
		Stomaphis quercus	–	AF163883[3]	–	AF156216[3]
		Tuberolachnus salignus	FM177113	FM174685	AJ298679[4]	–
	Tramini	*Protrama flavescens*	–	AF147814[3]	–	AF156206[3]
		Trama rara	–	AF147820[3]	–	AF156220[3]
Mindarinae		*Mindarus abietinus*	FM177107	FM174703	FM174681	–
Pterocommatinae		*Pterocomma pilosum*	AJ489283[19]	FM174701	AJ298672[4,j]	DQ005183[12,k]
Thelaxinae		*Thelaxes suberi*	AJ489287[19]	FM174702	AJ298667[4]	–
Family Phylloxeridae (outgroup)		*Daktulosphaira vitifoliae*	AJ489295[19]	FM174707	AJ298683[13]	DQ021446[13]

[a] Superscript numbers indicate sequences from previous studies (1) Stern (1994); (2) Moran et al. (1999); (3) Normark (2000); (4) Martínez-Torres et al. (2001); (5) von Dohlen et al. (2002); (6) Shingleton and Stern (2003); (7) von Dohlen and Teulon (2003); (8) Abbot and Withgott (2004); (9) Inbar et al. (2004); (10) Coeur d'acier et al. (2006); (11) von Dohlen et al. (2006); (12) Gao et al. (2000); (13) Baumann and Baumann (2005); (14) Stern et al. (1997); (15) Funk et al. (2000); (16) Rouhbakhsh et al. (1996); (17) Sunnucks and Hales (1996); (18) Clements et al. (2000); (19) Ortiz-Rivas et al. (2004). Superscript letters indicate sequences from a different species in the same genus: (a) *P. obscurus*, (b) *Geoica* sp., (c) *P. microsetosus*, (d) *T. populi-monilis*, (e) *C. bambusifoliae*, (f) *C. cupresi*, (g) *C. pinea*, (h) *Eulachnus* sp., (i) *L. shiicola*, (j) *P. populeum*, and (k) *P. populifoliae*.

rect sequencing of PCR products, using primers efs1, efr3 and internal primers efs2 (5′AAGGCTGAACGTGAACGTGGTATCAC3′) and efr4 (5′ATTTGACCNGGGTGRTTCAATAC3′). Sequencing was conducted using the Big Dye Terminator v3.1 Cycle Sequencing Kit

(Applied Biosystems) following manufacturer's instructions, and loading samples onto an ABI 3700 automated sequencer. All sequences of the COII gene and some sequences of the rest of genes were originally obtained in previous studies and retrieved from public databases for their analysis in the present report (see Table 1).

2.4. Computer analysis of DNA sequences

Chromatograms were analysed and assembled using the Staden package v1.6.0 (Staden et al., 1998). Multiple alignments were done with Clustal X v1.81 (Thompson et al., 1997) with gap opening and gap extension penalties of 10.0 and 0.2, respectively, and subsequently manually revised. Positions of introns were determined by aligning sequences obtained for genomic DNA with sequences obtained for cDNA of the species *Acyrthosiphon pisum* and *Cinara tujafilina* for the LWO gene and *Eriosoma lanuginosum*, *Pemphigus populi* and *Eulachnus rileyi* for the EF1α gene. Aligned amino acid sequences were obtained from aligned coding sequences using the translate option in MEGA version 3.1 (Kumar et al., 2004).

2.5. Phylogenetic analysis

Sequences were used both for separate analyses of each gene and for subsequent combined analyses. The number of species used in the separate analyses differed depending on the availability of sequences for each gene (34 for LWO, 46 for EF1α, 36 for ATP6, and 48 for COII, including outgroup sequences). Two sets of combined molecular data were created, concatenating the sequences of the different genes belonging to a same species or, in some instances, belonging to different species from the same genus (see Table 1). The first Combined Data Set (hereafter CDS) included the coding sequences of the nuclear LWO and EF1α genes and the mitochondrial ATP6 gene, from a total of 34 aphid species (or genera), representative of 11 subfamilies and the outgroup. In this CDS, third codon positions were removed from the LWO and ATP6 sequences because of saturation of nucleotide substitutions already revealed in previous studies (Martínez-Torres et al., 2001; Ortiz-Rivas et al., 2004), but all codon positions of the EF1α gene were included. A second Combined Data Set, named Total Evidence Set (hereafter TES), was created concatenating the total evidence from molecular data that could be compiled for 20 aphid species, representative of 6 subfamilies and the outgroup. The TES included all codon positions of the nuclear LWO and EF1α genes and the mitochondrial ATP6 and COII genes.

The program RRTree (Robinson-Rechavi and Huchon, 2000) was used to carry out global relative rate tests for each gene, as well as for the CDS, grouping the species in their respective subfamilies and tribes for the analysis. Additionally, specific comparisons were also carried out with PHYLTEST (Kumar, 1995).

Phylogenetic analyses of individual genes were carried out with maximum likelihood (ML) as implemented in PAUP* v4.0b10 (Swofford, 2002). Modeltest v3.7 (Posada and Crandall, 1998) was used in combination with PAUP* to find the model of DNA substitution that best fit each set of data, both in the separate and the combined analyses. TREE-PUZZLE v5.2 (Schmidt et al., 2002) was used to evaluate the homogeneity in composition of both nucleotide and aminoacid sequences. Phylogenetic reconstructions were inferred from the CDS and TES using ML, neighbour joining (NJ) and maximum parsimony (MP) in PAUP*, as well as Bayesian inference (BI) with MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001). ML searches were always performed with TBR branch swapping and 100 repetitions of random sequence addition, limiting rearrangements to 3000. MP heuristic searches were performed with TBR branch swapping and 5000 repetitions of random sequence addition. Distance based reconstructions were carried out with the NJ algorithm (Saitou and Nei, 1987) with distances calculated with the corresponding ML parameters obtained with Modeltest. Bootstrap (Felsenstein, 1985) was conducted for the three methods performing 300 replicates in ML and 2000 replicates in MP and NJ analysis. BI was also performed on the CDS and TES establishing a partition for each gene (3 and 4, respectively). Two parallel runs of six Markov chains were implemented, with 10^6 to 4×10^6 generations, until convergence between runs was reached, as checked with the program Tracer v1.4 (Rambaut and Drummond, 2003). Trees were saved every 100 generations and a consensus tree with posterior probabilities was calculated excluding a burn-in initial fraction of 25%.

Phylogenetic reconstructions were also carried out from the inferred aminoacid sequences of the genes, both separately and combining them in two data sets analogous to those established for nucleotide data. Aminoacid sequences of each gene were analysed by ML with PHYML (Guindon and Gascuel, 2003), using the aminoacid substitution model selected by ProtTest v2.4.4 (Abascal et al., 2005), with 300 bootstrap replicates. The ML analyses of the aminoacid CDS (that included the aminoacid sequences of the LWO, EF1α and ATP6 genes) and the TES (which also included the aminoacid sequences of the COII gene) were done using the same conditions. These data sets were also analysed by MP in PAUP*, with 100 repetitions of random sequence addition and 2000 bootstrap replicates, and with NJ in MEGA using the JTT model (which was always in a high position of the model rankings obtained in ProtTest) and 300 bootstrap replicates. They were also analysed with BI using MrBayes, partitioning the data into the 3 or 4 proteins, respectively, with the same conditions used in the BI from nucleotide sequences, but with just 5×10^5 generations, which were always enough for reaching convergence between the two parallel runs. The model of aminoacid substitution used for BI was MtREV + I + G (which, among the models available in MrBayes, was the one that occupied a higher position in the rankings of models that best fit the data obtained using ProtTest).

2.6. Tests for comparison of alternative phylogenetic hypotheses

SH tests (Shimodaira and Hasegawa, 1999) as implemented in PAUP* and ELW tests (Strimmer and Rambaut, 2002) as implemented in TREE-PUZZLE were conducted for the statistical comparison of alternative topologies, only on the nucleotide CDS. A total of 30 different topologies were compared at a time for each test, including ML topologies for each gene in the separate analyses, topologies for the combined analyses and for each reconstruction method, as well as topologies from Heie's and Wojciechowski's hypotheses from morphological data. Other topologies included ML trees constrained to keep some specific relationships proposed in the morphological hypotheses, as the sister relationship between the subfamilies Lachninae, Aphidinae and Pterocommatinae, or between Thelaxinae and the subfamilies Calaphidinae, Chaitophorinae and Drepanosiphinae (which were included in the family Drepanosiphidae in the classification proposed by Heie (1980)).

3. Results

3.1. Characterization of molecular variation

Table 2 shows a summary of data relative to the sequences analysed. Numbers of variable and parsimony informative positions for the EF1α gene were notably lower than those in the rest of genes, supporting the low evolutionary rates already described for this sequence in aphids (von Dohlen et al., 2002, 2006). The

Table 2
Summary of data relative to the genes analysed in this work and the Combined Data Sets.

Genes/ data set	Number of species[a]	Length of alignment[b]	Variable positions[c]	Informative positions[d]	Variable positions in aminoacids	Informative positions in aminoacids	Average *p*-distance[e]	Saturation in 3rd position[f]
LWO	34	739	165/383	124/322	103	75	0.168	Yes
EF1α	46	831	39/273	24/229	25	11	0.095	No
ATP6	36	651	196/399	137/325	131	108	0.191	Yes
COII	48	531	103/261	72/210	75	55	0.128	Yes
CDS	34	1757	620	473	257	193	–	–
TES	20	2749	1122	835	276	196	–	–

[a] Number of aphid species analysed, including the outgroup.
[b] Total number of nucleotide positions used in the phylogenetic analyses, including all codon positions of the genes except in the CDS, where third codon positions of LWO and ATP6 were excluded.
[c] Number of variable positions in 1st plus 2nd codon positions/number of total variable positions (excluding outgroup).
[d] Number of parsimony informative positions in 1st plus 2nd codon positions/number of total parsimony informative positions (excluding outgroup). For the CDS and TES, only total numbers of variable and parsimony informative positions are shown.
[e] Average *p*-distances among aphids, calculated excluding the outgroup, expressed in number of substitutions per nucleotide position.
[f] Evidence of saturation of nucleotide substitutions in third codon positions of the genes (see text).

mean *p*-distance among aphids for this gene (0.095) was also quite lower than that of the LWO gene (0.168).

Pairwise *p*-distances among aphid species were plotted against Kimura-2-parameter distances for the four sequences analysed (not shown). As previously reported, the LWO and ATP6 genes showed a marked saturation effect for third codon positions (Martínez-Torres et al., 2001; Ortiz-Rivas et al., 2004), which led us to exclude them in some of the analyses and especially from the Combined Data Set (CDS). No clear saturation effect was observed for third codon position of the EF1α gene. Besides, this position contained 85% of the variability found in the coding region of this gene (see Table 2), so all positions of the EF1α gene were included in the CDS. The plotting was also made for the COII gene, showing again a saturation effect for third codon positions. Despite the saturation observed for some of the genes, all codon positions of the four sequences were included in the analysis of the Total Evidence Set (TES).

The aminoacid composition of the four genes and the nucleotide composition of the mitochondrial genes was homogenous among species, as revealed by TREE-PUZZLE. On the contrary, nucleotide composition of the LWO gene was not homogenous because of a higher G + C content observed in most Fordini species (Eriosomatinae) and in *Tuberolachnus salignus* (Lachninae: Lachnini) and a higher A + T content observed both in *Tetraneura caerulescens* (Eriosomatinae: Eriosomatini) and the outgroup species *Daktulosphaira vitifoliae*. For the EF1α nucleotide sequences, a bias was also observed in *Baizongia pistaciae* (Eriosomatinae: Fordini) towards higher G + C composition, and in *D. vitifoliae* towards higher A + T content. Notably, the species *T. caerulescens* also coincided with the outgroup in a higher A + T content for the EF1α sequence, although it was not statistically significant ($p = 0.091$).

3.2. Relative rates tests

Global relative rates tests implemented on RRTree and subsequent pairwise relative rates tests on PHYLTEST allowed the detection of several groups with significantly accelerated rates of nucleotide and/or aminoacid substitution (see Supplementary Tables 1 and 2). It is worth mentioning the accelerated rates detected for the tribes Fordini (Eriosomatinae) and Lachnini (Lachninae) and *Cerataphis* sp. (Hormaphidinae) for most of the genes analysed, as well as for the nucleotide CDS. *Mindarus abietinus* (Mindarinae) also showed significant accelerated rates for the LWO gene and, especially, the EF1α gene in some of the comparisons, and also when the nucleotide CDS was used. The species of the tribe Eriosomatini also showed accelerated rates when compared to the tribe Fordini for the ATP6 gene and the aminoacid CDS. Finally, the subfamily Calaphidinae displayed higher rates for nucleotide sequences of the mitochondrial ATP6 gene and also for both nucleotide and aminoacid sequences of the LWO and the CDS. No group displayed an accelerated relative evolutionary rate for the mitochondrial COII gene, either considering nucleotide or aminoacid data.

3.3. Phylogenetic reconstruction from separate analyses of the genes

Despite the clear saturation of third codon positions detected for some of the genes, the phylogenetic relationships obtained in the separate analyses when these positions were included were roughly similar to those excluding them. Nevertheless, the statistical support to most of the nodes was much lower when only first and second codon positions were analysed, especially for the mitochondrial genes. The phylogenetic reconstructions obtained in the separate analyses of the four genes showed in general the grouping of the species in their respective tribes according to classifications. The monophyly of tribes was more general and supported by higher bootstrap values when the analyses were carried out on nuclear sequences (Figs. 2 and 3) than on mitochondrial sequences (see Supplementary Figs. 1 and 2). On the contrary, most of the topologies obtained in the separate analyses showed very little support for phylogenetic structure for taxonomic levels higher than tribe, similarly to the results obtained in previous reports (von Dohlen and Moran, 2000; Martínez-Torres et al., 2001). The support for the monophyly of the subfamilies analysed was highly variable across the separate analyses. Almost all the subfamilies analysed grouped as monophyletic in some of the topologies, but were non-monophyletic in others. Only the subfamily Chaitophorinae always grouped as monophyletic. The monophyly of the subfamily Hormaphidinae appeared in some reconstructions obtained in the analyses of the EF1α gene (Fig. 2) and when third codon positions were excluded from the COII gene (but not when they were included, see Supplementary Fig. 2), although with a low associated statistical support, and always a closer relationship between the tribes Hormaphidini and Nipponaphidini was observed (see Fig. 2). The subfamily Aphidinae was not monophyletic in the reconstructions of the EF1α gene, as previously reported (von Dohlen et al., 2006), due to the sister relationship between the tribe Macrosiphini and the representative of the subfamily Pterocommatinae, *Pterocomma pilosum* (see Fig. 2). On the contrary, the analyses of the LWO, ATP6 and COII genes showed a sister relationship between a monophyletic Aphidinae and the representative of Pterocommatinae (see Fig. 3 and Supplementary Figs. 1 and 2). Some reconstructions highly supported the monophyly of the subfamilies Calaphidinae (see Figs. 2 and 3 and Supplementary Fig. 1)

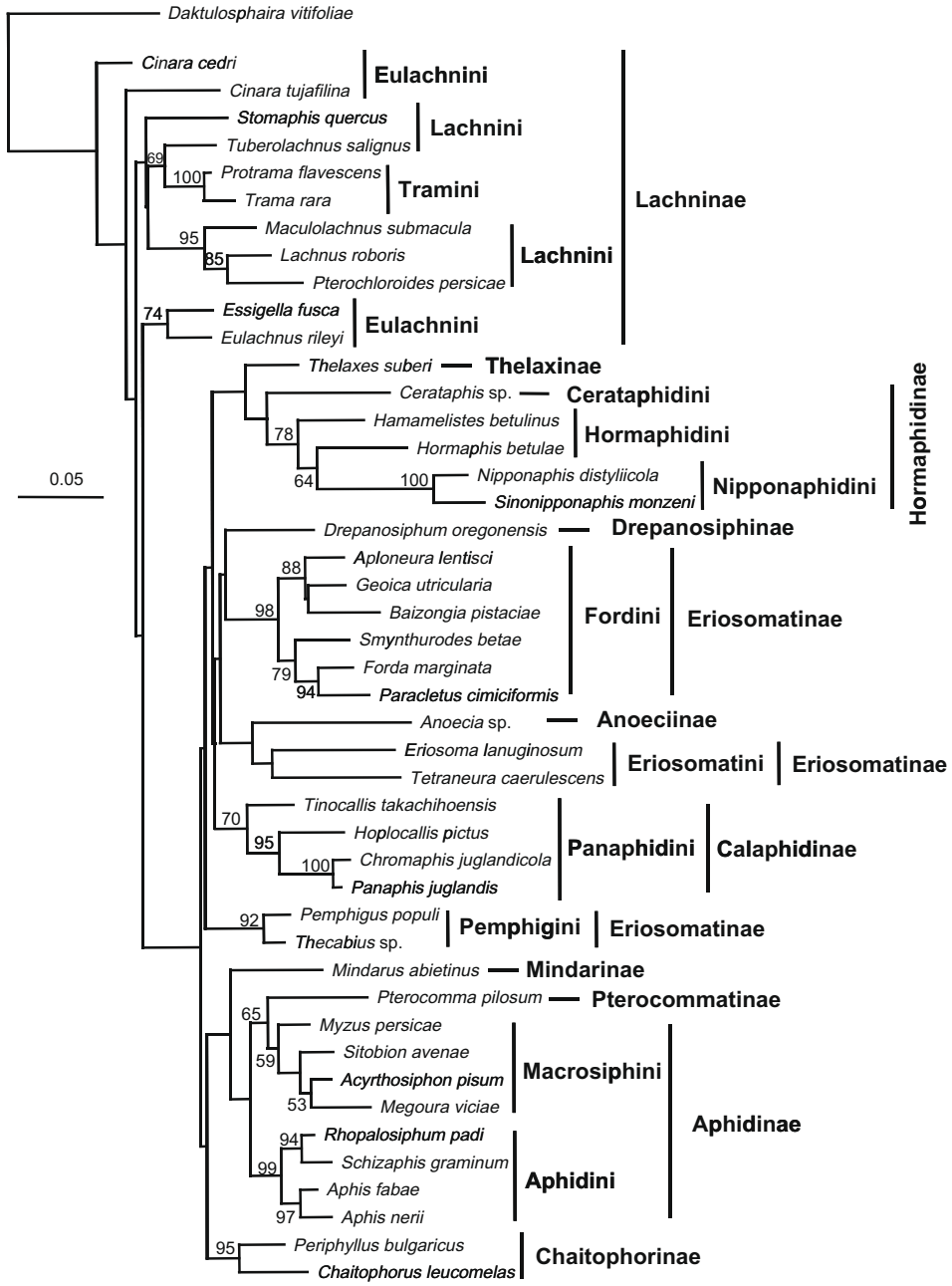

Fig. 2. ML tree obtained from the analysis of the coding region of the EF1α gene. Bootstrap values are shown above or below branches only when higher than 50%.

and Lachninae (see Fig. 3 and Supplementary Fig. 2), and some topologies also showed a monophyletic group compatible with the family Drepanosiphidae *sensu* Heie (1980), including in it the representatives of the subfamilies Calaphidinae, Chaitophorinae and Drepanosiphinae (see Fig. 3). The subfamily Eriosomatinae was almost always polyphyletic because the tribe Eriosomatini was usually placed in a distant position with respect to the tribes Fordini and Pemphigini. This subfamily was only monophyletic in the reconstructions obtained from the COII gene, for which no sequences for the tribe Eriosomatini are currently available.

The lack of resolution for the deepest nodes of the tree was especially noticeable for the nuclear EF1α gene (see Fig. 2) and the mitochondrial ATP6 and COII genes (see Supplementary Figs. 1 and 2). In contrast, the phylogenetic reconstructions obtained in the analysis of the LWO gene yielded greater support for the oldest relationships (see Fig. 3 and Supplementary Fig. 5). The addition of new species to the analysis of this gene did not significantly alter the results previously obtained (Ortiz-Rivas et al., 2004), except for the phylogenetic position of *Mindarus abietinus* (see below). The initial, rooted ML reconstruction placed the representatives of the subfamily Calaphidinae in the basal position, distantly related to other representatives of Heie's family Drepanosiphidae. The existence of accelerated rates and/or compositional biases detected for this subfamily and also for the tribe Eriosomatini (see above) could be provoking and attraction of these groups to the outgroup. For these reasons, phylogenetic reconstructions were also made excluding the subfamily Calaphidinae and the tribe Eriosomatini, on one hand, or excluding the outgroup species, *Daktulosphaira vitifoliae*, on the other. The unrooted analysis of the LWO gene allowed the recognition of three main lineages in which the aphid subfamilies analysed could be grouped (see Fig. 3). Following Ortiz-Rivas et al. (2004) these lineages were named A + D

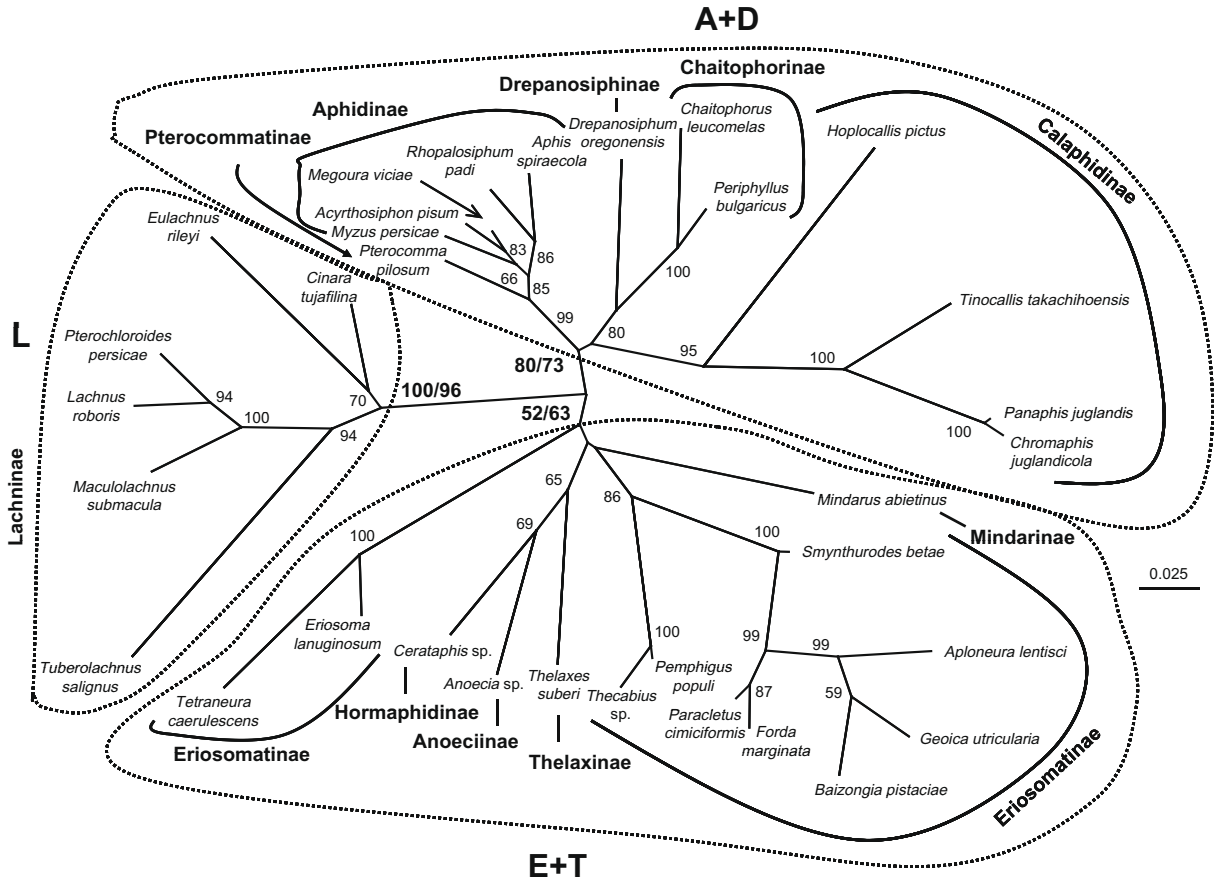

Fig. 3. Unrooted ML reconstruction obtained from the complete coding sequence of the LWO gene (third codon positions included). Dashed ellipses enclose the three main lineages in which the aphid subfamilies analysed are grouped, named as A + D, E + T and L (see text). Bootstrap values are shown next to nodes only when higher than 50%. For the three main lineages, bootstrap values from the nucleotide analysis are shown before the slash and from the aminoacid ML analysis after the slash.

(including the subfamilies Aphidinae, Calaphidinae, Chaitophorinae, Drepanosiphinae and Pterocommatinae), E + T (subfamilies Anoeciinae, Eriosomatinae, Hormaphidinae, Mindarinae and Thelaxinae) and L (composed of the representatives of the subfamily Lachninae). Bootstrap support for these three groupings ranged from low to very high (see Fig. 3). Moreover, excluding the Calaphidinae and the Eriosomatini from the unrooted analysis had the general effect of increasing the bootstrap support for the three lineages, which ranged from 88% to 100%. The position of the representative of the subfamily Mindarinae in the analysis of the LWO sequence, inside the lineage named E + T, differed from the position obtained in the reconstructions using the EF1α and ATP6 genes, where *Mindarus abietinus* branched as a sister group of the clade composed of Aphidinae + Pterocommatinae. Bootstrap values for the position of *Mindarus abietinus* for the two latter genes were, nevertheless, always very low.

Based on the results obtained in the unrooted analyses of the LWO gene, the phylogenetic reconstructions were done again including the outgroup but constraining the topologies to keep the three main lineages described, in order to analyse the relationships among them. The ML constraint analysis of the LWO sequences placed the lineage A + D in the basal position, both including third codon positions (68% of bootstrap support) or excluding them (55% of bootstrap support; not shown). On the contrary, the individual constraint analyses of the EF1α, ATP6, and COII genes always yielded topologies in which the subfamily Lachninae was sister to the rest of lineages, with up to 100% of bootstrap support in the EF1α constraint ML tree.

3.4. Phylogenetic reconstruction from the Combined Data Set and the Total Evidence Set

Most of the rooted and unrooted analyses obtained from the Combined Data Set (CDS), which included the first and second codon positions of the LWO and ATP6 genes and all codon positions of the EF1α gene for 34 aphid species, showed a topology almost coincidental with the unrooted LWO analysis. The three main lineages could be recognised except for the unstable position of the tribe Eriosomatini, which usually grouped with the subfamily Lachninae or was basal in the tree. The grouping of species in tribes was generally well supported but, contrary to the separate analyses, the monophyly of most of the subfamilies was commonly well supported too. The subfamilies Calaphidinae, Chaitophorinae and Lachninae generally appeared as monophyletic groups (Fig. 4). When the tribe Eriosomatini was included in the analyses, the subfamily Eriosomatinae was always polyphyletic because of the position of this tribe (results not shown). A monophyletic group compatible with Heie's Drepanosiphidae (see Fig. 4) was also present in most of the analyses. The subfamily Aphidinae was always monophyletic and showed a highly supported relationship with the subfamily Pterocommatinae.

A few reconstructions were obtained in which the topology showed exactly the three lineages previously described, with the Eriosomatini inside lineage E + T and without the need of a constraint analysis. This topology was obtained, for example, when the aminoacid CDS was analysed by NJ and the outgroup was included in the analysis. Some of the unrooted analyses of aminoacid data also

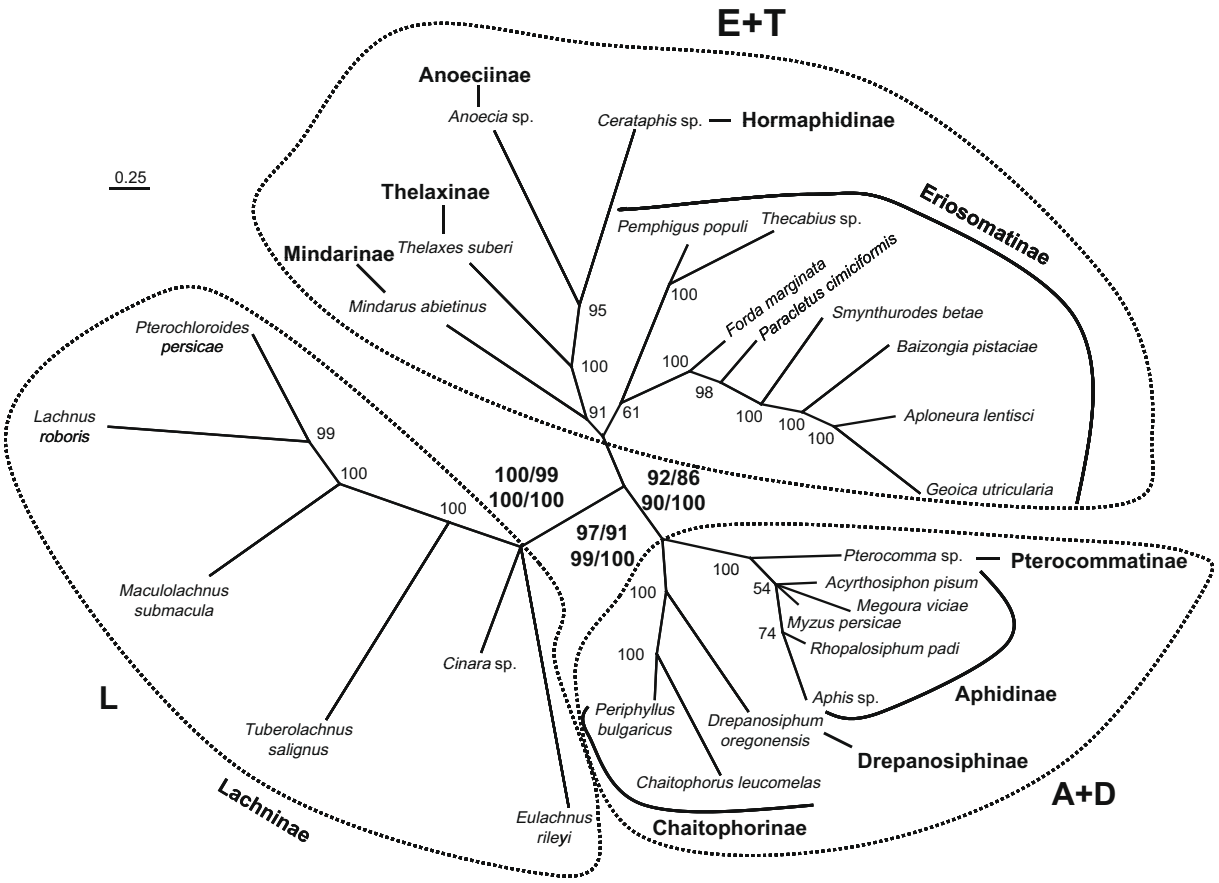

Fig. 4. Unrooted bayesian inference of aphid phylogeny resulting in the analysis of the Combined Data Set when the species of the tribe Eriosomatini and the subfamily Calaphidinae were excluded. Posterior probabilities are shown next to nodes. Dashed ellipses enclose the three main lineages A + D, E + T and L, for which several statistical supports are shown: ML bootstrap/MP bootstrap/NJ bootstrap/BI posteriors.

displayed the three main lineages topology, in which the three groups were supported by bootstrap values ranging from low to high. When the tribe Eriosomatini and the subfamily Calaphidinae, both affected by compositional biases and/or accelerated relative rates of evolution, were excluded from the unrooted analysis, the statistical support for the three main lineages was always very high, independently of the reconstruction method chosen (see Fig. 4).

In the analysis of the CDS, *Mindarus abietinus* was placed always inside the lineage E + T, supporting the results obtained in the separate analysis of the LWO gene and differing from the topologies of the EF1α and ATP6 genes. Some other important relationships found in the combined analyses and also in some of the separate analyses, include the sister relationships between the subfamilies Chaitophorinae and Drepanosiphinae (also present in the results obtained by von Dohlen and Moran (2000)), on one hand, and between the tribes Fordini and Pemphigini, belonging to the subfamily Eriosomatinae, on the other. It is worth mentioning the existence of a group composed of the representatives of the subfamilies Anoeciinae, Hormaphidinae and Thelaxinae that showed a 100% posterior probability in the BI carried out on the aminoacid CDS (see Fig. 4).

A constraint analysis forced to keep the three main lineages previously described was also carried out on the nucleotide and aminoacid CDS. All the methods used yielded topologies in which the lineage named L, composed of the representatives of the subfamily Lachninae, was placed in the basal position of the tree, with moderate to high values of statistical support (Fig. 5).

The analysis of the Total Evidence Set (TES) included all codon positions of the four genes used in this study, on one hand, and of the corresponding inferred aminoacid sequences, on the other. Although only 20 aphid species or genera could be included in the analyses, most of the topologies obtained were compatible with the three main lineages described above (see Supplementary Fig. 3). Furthermore, some of the unconstraint topologies also supported a basal position for the subfamily Lachninae, although with low or moderate statistical support.

3.5. Tests for comparison of alternative phylogenetic hypotheses

Table 3 shows the results of the SH and ELW tests performed on a set of 30 alternative topologies using as data the nucleotide CDS. ELW tests yielded a confidence set that included only 10 topologies, excluding Heie's and Wojciechowski's proposals, shown in Fig. 1. The confidence set also excluded most of the topologies obtained after nucleotide separate analyses and topologies from separate and combined aminoacids analyses. The topologies that were included to test the sister relationship between some subfamilies, following the morphological phylogenetic proposals, were also excluded from the confidence set. Particularly, the topologies forced to keep the relationship among Lachninae, Aphidinae and Pterocommatinae or among Thelaxinae and Heie's Drepanosiphidae were excluded. The set also excluded the topologies constrained to keep the three main lineages and either the lineage A + D or E + T in the basal position, but included that with the subfamily Lachninae in this position. SH test proved to be much more conservative than ELW test, only rejecting some of the topologies from separate analyses of nucleotide and aminoacid data. Aware that SH test are sensitive to the inclusion of very unlikely topolo-

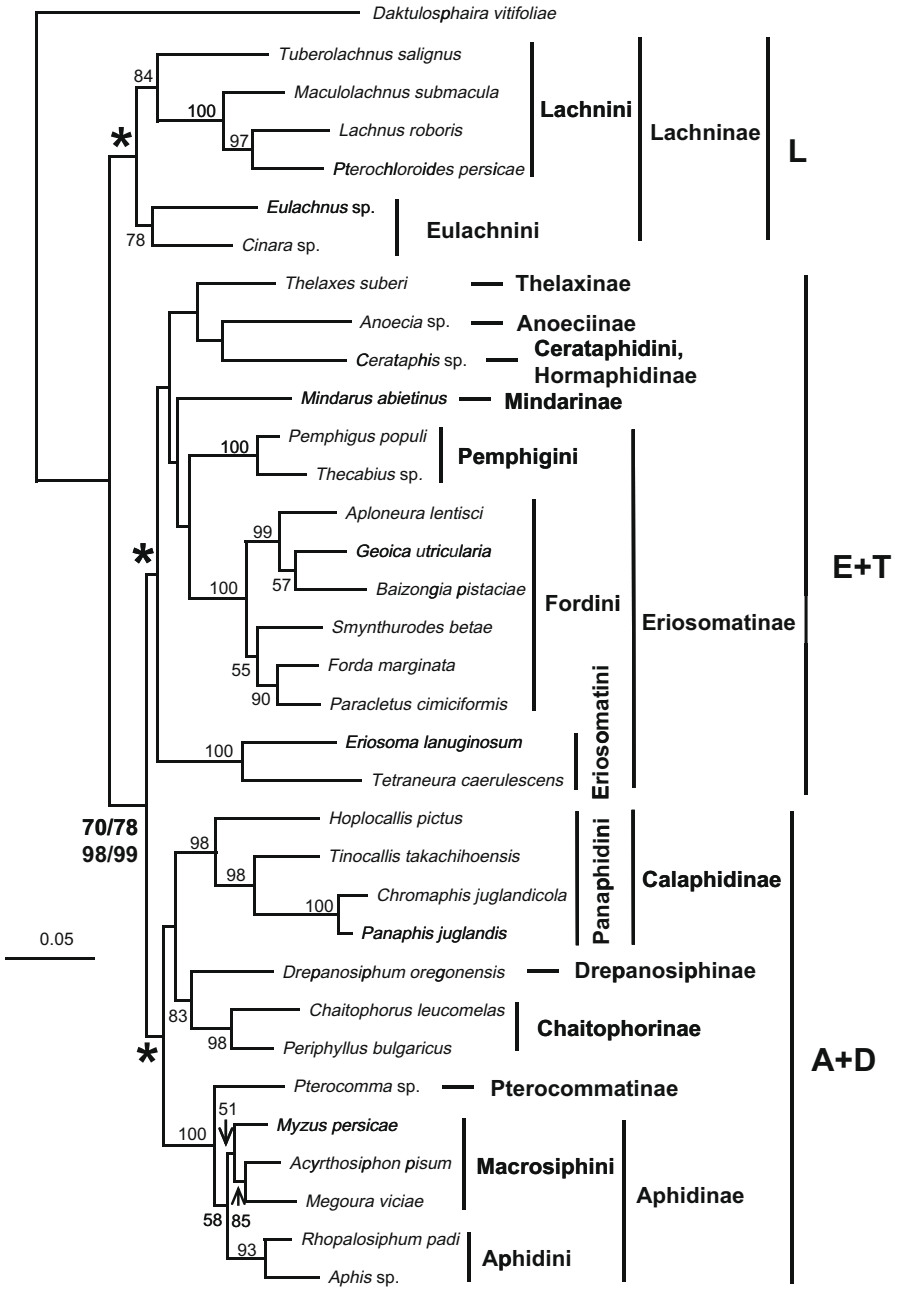

Fig. 5. Phylogenetic tree resulting from the ML analysis of the nucleotide Combined Data Set when the topology was constrained to keep the three main lineages A + D, E + T and L. An asterisk (*) marks those nodes that were constrained to appear. For the different methods used, the relationships among the three lineages were always solved with the basal position of the lineage L, composed of the subfamily Lachninae. For this node, different statistical values are shown: ML bootstrap/MP bootstrap/NJ bootstrap/BI posteriors.

gies (Strimmer and Rambaut, 2002) we performed the test on a set of 25 topologies that excluded the 5 ones that were rejected in the first set. In this second set, only Heie's topology was significantly worse, with Wojciechowski's topology near rejection with a *p*-value = 0.082.

4. Discussion

The first attempts to infer the phylogeny of aphids with molecular data bumped into a striking lack of phylogenetic structure for taxon levels higher than tribe (von Dohlen and Moran, 2000; Martínez-Torres et al., 2001). This lack of resolution power for the deepest nodes of the tree of the mitochondrial genes used was interpreted as a consequence of the rapid adaptive radiation that aphids are thought to have undergone during the Late Cretaceous, when most of the extant subfamilies probably appeared (Heie, 1987, 2004; von Dohlen and Moran, 2000; Martínez-Torres et al., 2001). Nevertheless, the subsequent use of the nuclear gene encoding the long-wavelength sensitive opsin (LWO) proved that molecular data can be compiled to obtain enough information for eventually reaching a good resolution of the tree of the family Aphididae (Ortiz-Rivas et al., 2004). The analysis of this gene indicated the existence of three main lineages of aphids, and suggested a basal position of the subfamily Lachninae in the phylogeny of the family Aphididae.

Table 3
Results from SH and ELW tests performed on a set of 30 alternative topologies.

	Breaf description of topology	ELW		SH	
		δ	c	δ	p-Value
1	CDS ML tree	0.13	0.1617	Best	Best
2	CDS constrained ML tree (basality of lineage L; Fig. 5)	7.18	0.0174*	7.64	0.974
3	Heie's topology	77.22	0.0000*	74.31	0.072
4	Wojciechowski's topology	53.43	0.0000*	50365	0.295
5	LWO ML tree	7.82	0.0383*	9.14	0.934
6	LWO ML tree 1st and 2nd positions only	58.29	0.0030*	58.51	0.221*
7	EF1α ML tree	129.58	0.0000*	139.25	0.008*
8	ATP6 ML tree	217.65	0.0000*	218.34	0.000*
9	ATP6 ML tree 1st and 2nd positions only	151.01	0.0000*	150.34	0.004*
10	CDS 2nd ML tree	Best	0.2114*	2.06	0.975
11	LWO ML tree constrained to lineages A + D, E + T and L	13.53	0.0038*	14.38	0.867
12	CDS ML tree constrained to basality of Eriosomatini	2.36	0.0611	3.25	0.994
13	CDS ML tree constrained to basality of species of E + T (not as a lineage)	2.26	0.1275*	5.84	0.950
14	CDS ML tree constrained to Thelaxinae + Heie's Drepanosiphidae	45.70	0.0001*	44.80	0.393
15	CDS ML tree constrained to Aphidinae + Lachninae	15.37	0.0110*	16.31	0.839
16	CDS ML tree constrained to Mindarinae + Aphidinae	20.12	0.0394*	21.97	0.788
17	CDS ML tree constrained to Mindarinae + Heie's Drepanosiphidae	35.46	0.0000*	36.90	0.513
18	CDS ML tree constrained to Mindarinae + Thelaxinae + Heie's Drep.	46.15	0.0009*	48.85	0.312
19	CDS ML tree constrained to monophyly of Eriosomatinae	7.82	0.0459	9.44	0.920
20	CDS MP tree	2.36	0.1750	7.02	0.919
21	CDS BI tree	4.68	0.0755*	5.93	0.956
22	CDS NJ tree	25.21	0.0022*	26.14	0.726
23	Aminoacid CDS ML tree	33.20	0.0076*	33.12	0.596
24	Aminoacid CDS MP tree	25.49	0.0131*	28.26	0.663
25	Aminoacid CDS BI tree	37.87	0.0017*	37.89	0.515
26	Aminoacid CDS NJ tree	30.94	0.0010*	31.80	0.611*
27	LWO aminoacids ML tree	203.50	0.0000*	198.65	0.000*
28	ATP6 aminoacids ML tree	154.98	0.0000*	156.86	0.001*
29	CDS ML tree constrained to basality of lineage E + T	13.72	0.0013*	13.50	0.895
30	CDS ML tree constrained to basality of lineage A + D	26.88	0.0012*	26.12	0.690

The complete set of topologies can be found in Supplementary Fig. 4. CDS, Combined Data Set; δ, difference in $-\ln L$ from best topology as calculated by the programs; c, confidence value (expected likelihood weight). A $*$ symbol denotes topologies significantly out of the confidence set in the ELW test and topologies significantly worse than the ML tree in the SH test.

In this report we have presented the results of a broader survey of aphid species and the compilation of a bigger set of molecular data, which gives further support to our previous results. The phylogenetic reconstructions obtained support the grouping of the subfamilies analysed in three main lineages (Fig. 6), which were named A + D: (subfamilies Aphidinae, Calaphidinae, Chaitophorinae, Drepanosiphinae and Pterocommatinae), E + T (subfamilies Anoeciinae, Eriosomatinae, Hormaphidinae, Mindarinae and Thelaxinae) and L (subfamily Lachninae).

4.1. Support for the existence of three main lineages in the family Aphididae

The separate analyses of the nuclear elongation factor 1α gene (EF1α) and the mitochondrial genes encoding the subunit 6 of the F-ATPase (ATP6) and the subunit II of the cytochrome oxidase (COII) have shown again the difficulty of obtaining a highly resolved topology for the phylogenetic tree of aphids, especially for the deepest nodes. The only exception to date for this lack of resolution power for single sequence analyses has come from the use of the nuclear LWO gene, which highly supported some relationships among the subfamilies of aphids analysed and especially the existence of the three lineages described above. Despite the little phylogenetic structure found in the separate analyses of the EF1α, ATP6 and COII genes, their combination with the LWO gene had the general effect of increasing the support to the monophyly of several subfamilies. The analysis of the Combined Data Set (CDS) further supported the existence of the three main lineages, which displayed high statistical values, especially when the species of the tribe Eriosomatini and the subfamily Calaphidinae were not included (see Fig. 4). Furthermore, all the methods used to analyse the CDS supported a basal position of the subfamily Lachninae in the phylogeny of aphids (see Fig. 5).

A series of morphological characters can be invoked to support the existence of three main lineages in the family Aphididae. Each of these three lineages show a group of characteristics that unite its components and distinguish them from the other lineages (see Fig. 6). Wojciechowski (1992) defined two "developmental lines" in aphids depending on the form of the gut and the presence or absence of triommatidium in the first instar larvae. One of the developmental lines would be characterized by a short gut (apomorphic state according to Wojciechowski) and the presence of triommatidium in first instar larvae (plesiomorphic according to this author). This line would be composed of the subfamilies Anoeciinae, Eriosomatinae, Hormaphidinae, Mindarinae and Thelaxinae (plus two subfamilies not represented in the present work: Phloeomyzinae and Greenideinae). Our results usually grouped these subfamilies inside the lineage that we have named E + T. The first three subfamilies were also grouped together by Heie, but he joined Mindarinae and Thelaxinae to his Drepanosiphidae (Heie, 1987) (see Fig. 1). These two subfamilies differ from the rest of members of the lineage E + T in their life cycles, because they display monoecious cycles. The subfamilies Anoeciinae, Eriosomatinae and Hormaphidinae show dioecious cycles with a characteristic "pemphigid" type of alternation (with a morph called sexupara returning to the primary host).

The separate phylogenetic analyses of the genes used in this study have shown a striking incongruence in the position of *Mindarus abietinus*, the only representative of the subfamily Mindarinae used in this report. This species clustered next to the subfamilies Aphidinae and Pterocommatinae in the analyses of the EF1α and ATP6 genes, but appeared inside the lineage E + T in the analysis of the LWO gene. Despite these contradictory positions, this species branched consistently inside lineage E + T in the combined analysis. This result agrees with Wojciechowski's proposal, where this subfamily is placed as a sister taxa of a group composed of

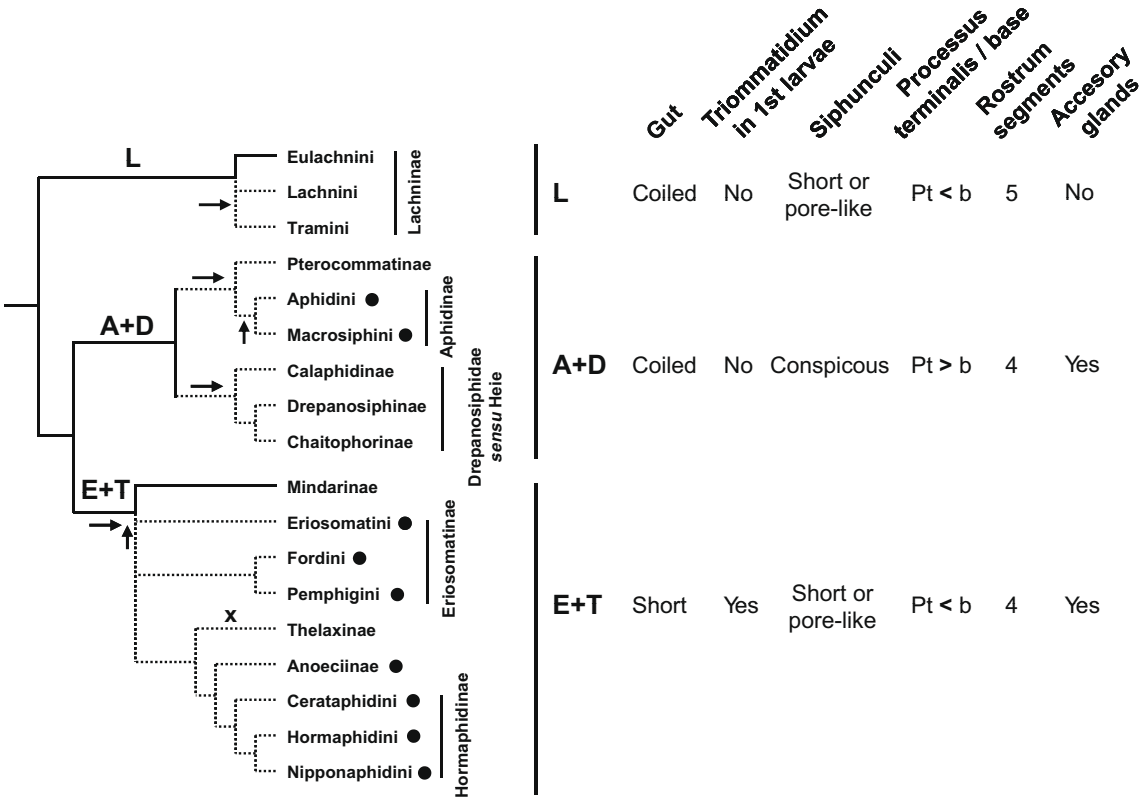

Fig. 6. A proposal for the phylogenetic relationships among the aphid subfamilies analysed in this study. The phylogenetic tree represents a consensus of the analyses carried out, with the three main lineages described and the subfamily Lachninae occupying the basal position. Black lines represent aphid groups feeding on gymnosperms and dotted lines aphid groups feeding on angiosperms. Filled circles next to taxa names indicate those groups with host alternating cycles. Depending on the true relationships inside lineage E + T, the proposal might support an evolutionary scenario with four transitions of feeding from gymnosperms to angiosperms (horizontal arrows) and a minimum of just two origins of host alternating cycles (vertical arrows), which would imply the loss of this character in Thelaxinae (marked with a cross). On the right, the state of a series of morphological characters in each of the three main lineages is shown (extracted from Wojciechowski, 1992; see text).

the subfamilies Anoeciinae, Eriosomatinae, Hormaphidinae and Thelaxinae based on the short gut and the presence of triommatidium in the first instar larvae, which characterises this "developmental line". The incongruence in the position of *M. abietinus* in the separate analyses might be due to the existence of accelerated rates in this species for some of the genes (see Section 3).

The second "developmental line" proposed by Wojciechowski would be characterized by a coiled gut and the absence of triommatidium in the first instar larvae. This line would include the subfamilies Aphidinae, Calaphidinae, Chaitophorinae, Drepanosiphinae, Lachninae and Pterocommatinae (plus other subfamilies not represented in the present work and that would be included in the family Drepanosiphidae *sensu* Heie). Our results support grouping all these subfamilies together except Lachninae, in the lineage we named A + D. All the subfamilies included in this lineage display two characters that distinguish them from the other lineages. Most of the representatives of this lineage show well developed siphunculi, as mentioned above, and the processus terminalis of the last segment of the antennae is longer than 0.5 times its base. In the lineages L and E + T, siphunculi are short or even pore-like, and the processus terminalis is shorter than 0.5 times its base. The relative length of the two parts of the last segment of the antennae (base and processus terminalis) seems to be a good diagnostic character, as its state is stable in higher taxa, according to Wojciechowski (1992).

The results obtained in this work from molecular data greatly disagree with Heie's proposal for the phylogeny of aphids, but are more similar to Wojciechowski's topology with respect to the relationships among major groups (see Figs. 1 and 6). The main disagreement concerns the position of the subfamily Lachninae, also named here as lineage L, which was thought to be the sister subfamily of Aphidinae + Pterocommatinae in both morphological proposals. However, our results do not agree with this hypothesis and suggest that Lachninae is the basal group of the family Aphididae. Not only combined data but most of the separate analyses of genes in this and previous works place Lachninae far from Aphidinae + Pterocommatinae, and the tests for comparisons of alternative topologies carried out in the present study rejected a topology constrained to keep this relationship (see Section 3). Besides, the topologies constrained to keep the three main lineages were also rejected except the one with Lachninae basal to the rest of aphids. Several characters distinguish aphids belonging to this subfamily from aphids belonging to the other two lineages (see Fig. 6), including five segmented rostrum (four in the rest of subfamilies) and lack of accessory glands in the male reproductive system (present in the rest) (Wojciechowski, 1992). According to Heie (1987), the state of many of the characters in the subfamily Lachninae are plesiomorphic or seem to be, and the age that aphidologist have given to this group depends on wether they considered these states as plesiomorphic or not.

4.2. Implications on aphid life cycles

A clear picture of the evolution of aphid life cycles is hampered by the lack of resolution for some of the nodes in the results presented in this report, especially inside lineages E + T and L. Concerning the evolution of feeding in aphids, the most parsimonious explanation would be accepting that the ancestor of extant aphid species lived on angiosperms and that, for the subfamilies analysed, only two transitions of feeding from angiosperms to gimnosperms occurred, one for the tribe Eulachnini

and another for the subfamily Mindarinae. Nevertheless, several authors have proposed that the ancestral life cycle of viviparous aphids was a non-alternating cycle on conifers (Heie, 1987; Shaposhnikov, 1987). Besides, aphidologists claim that the current feeding on conifers of Mindarinae, and also of Neophyllaphidinae (not included in this report) are ancestral for these two subfamilies. Assuming an ancestral feeding on conifers, a possible scenario for the evolution of feeding with 4 transitions from gymnosperms to angiosperms could be proposed for the subfamilies analysed (see Fig. 6). A first transition could have occurred in the ancestor of the tribes Lachnini and Tramini. However, our results do not clearly support a close relationship between this two tribes. Only some species of the tribe Lachnini seem to be clearly more related to the Tramini than to the Eulachnini and the rest of Lachnini (see for instance *Tuberolachnus salignus* in Fig. 2 and *Nippolachnus piri* in Supplementary Fig. 2; see also Normark, 2000). For the lineage A + D the most parsimonious explanation would imply just one transition, but we find it necessary to propose two, one for the ancestor of Aphidinae + Pterocommatinae and another one for the ancestor of Calaphidinae + Chaitophorinae + Drepanosiphinae (see Fig. 6). This is because the subfamily Neophyllaphidinae, with a likely ancestral feeding on conifers, was also a component of the family Drepanosiphidae *sensu* Heie, like the subfamilies Calaphidinae, Chaitophorinae and Drepanosiphinae. For the lineage E + T, the uncertainty about the relationships among its subfamilies, and especially about the position of Mindarinae, makes it difficult to discuss about the evolution of life cycles in the group. However, if this subfamily (the only one in this lineage that currently feeds on gymnosperms) was basal in this group (as proposed by Wojciechowski; see Fig. 1), then the most parsimonious explanation would imply just one transition from gymnosperms to angiosperms for the rest of subfamilies in this lineage (see Fig. 6).

The lack of resolution for higher taxonomical levels and the long history of aphid tribes found by von Dohlen and Moran (2000) led these authors to propose the existence of seven independent origins of host alternating life cycles in these group of insects. Some of the relationships that were statistically supported in the present study should lead to the proposal of a lower number of origins. As von Dohlen and Moran pointed out, the different kind of host alternation found in the Aphidinae, with gynoparae and males returning independently, supports at least an independent origin of this feature for this subfamily. This fact is also supported by the distant phylogenetic position that the Aphidinae had in the topologies obtained in this report with respect of the rest of alternating groups. All these groups were included in the lineage E + T, and all of them display an alternating life cycle with a return to the primary host carried out by a sexuparae. Again, the lack of resolution for the relationships within this lineage impedes the proposal of a clear picture for the evolution of this trait. Nevertheless, the strong support found for the relationship of Fordini and Pemphigini suggests a unique origin of host alternation for both instead of two. The monophyly of the subfamily Hormaphidinae observed in some of the reconstructions obtained from the EF1α and COII sequences also suggests that only one origin could have occurred in the ancestor of this subfamily. However, a most parsimonious explanation could be again that Mindarinae was basal in this group and only one origin of host alternating life cycles would have occurred for all the rest of subfamilies of the lineage E + T. Nevertheless, this explanation would imply accepting the loss of this character in the subfamily Thelaxinae, characterized by monoecious cycles.

4.3. Implications on the coevolution between aphids and Buchnera

The hypothesis of the strict coevolution between aphids and their primary endosymbiont *Buchnera aphidicola* is based on the way this bacterium is transmitted from mothers to daughters and in the coincidence between the first molecular phylogenies obtained from *Buchnera* sequences and those of their respective aphid hosts following Heie's proposal. Nevertheless, these molecular phylogenies included primary endosymbiont sequences of a very limited number of aphid subfamilies. To date, only one study approached the phylogeny of *Buchnera* with a good representation of subfamilies (Martínez-Torres et al., 2001). The topologies obtained in that report showed some disagreements with Heie's phylogeny, but were compatible with the molecular phylogenies obtained for aphids in the previous and the present work. However, the lack of resolution that also affected the deepest nodes of *Buchnera* phylogenetic reconstructions hampers the comparison with the results obtained in the present study for their hosts, the aphids. Nevertheless, some specific relationships can be compared, among which it is worth mentioning that several phylogenetic reconstruction made using *Buchnera* DNA sequences placed the endosymbionts of the subfamily Lachninae in the basal position of the trees. For example, this position was supported by a 91% of bootstrap support in the MP analysis of the aminoacid sequences of the gene enconding the subunit β of the F-ATPase complex (Martínez-Torres et al., 2001). Interestingly, genome size of *Buchnera* belonging to species of the Lachninae are the smallest ones found to date on aphids, which is in agreement with a longer independent evolutionary history of this subfamily (Gil et al., 2002; Perez-Brocal et al., 2006). A molecular phylogeny of *Buchnera* with both a good representation of aphid subfamilies and a high resolution for the oldest relationships is still needed to definitely confirm the hypothesis of the strict parallel evolution among these two taxa.

4.4. Perspectives on aphid systematics and evolutionary studies

Despite the lack of phylogenetic structure for the deepest nodes of the phylogeny of aphids that was found in the first molecular studies using mitochondrial sequences, the subsequent use of the nuclear genes that encode the long-wavelength opsin and the elongation factor 1α has allowed important advances in the knowledge of the relationships among aphid subfamilies. Consequently, the sequencing of new nuclear genes should hopefully allow in the future the reconstruction of a robust phylogeny of the family Aphididae. At the same time, future studies should improve the sampling of aphid subfamilies, to definitely validate the existence of the three main lineages proposed or reveal new ones. The inclusion of representatives of the subfamily Neophyllaphidinae would be of great help to the discussion of the evolution of feeding in this group of insects. A better representation of subfamilies that were included in the family Drepanosiphidae *sensu* Heie (1980) would also allow to a test of its monophyly and the validity of its split in 14 subfamilies in the classification of the *Catalogue of the World's Aphididae* (Remaudière and Stroyan, 1984; Remaudière and Quednau, 1988; Quednau and Remaudière, 1994; Remaudière and Remaudière, 1997). Finally, our results that support the existence of three main evolutionary lineages in the Aphididae are also relevant for the current genomic studies on aphids. If additional genomes are to be sequenced for evolutionary comparative purposes, then the topologies obtained in the present work would suggest choosing a representative of the lineage E + T and another of the lineage L. This objective, added to the recent sequencing of the genome of the pea aphid *Acyrthosiphon pisum*, would provide one representative genome from each of the three main aphid lineages described in this report.

Acknowledgments

The authors wish to thank J.M. Michelena, P. González and N. Pérez Hidalgo for providing some and identifying most of the specimens used in this work. This work was supported by the Spanish MEC with project CGL20007-67392. B.O.R. enjoyed a Ph.D. fellow-

ship from the Government of Generalitat Valenciana, Spain. The facilities at SCSIE (Universitat de València) were used for sequencing.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.12.005.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Abbot, P., Withgott, J.H., 2004. Phylogenetic and molecular evidence for allochronic speciation in gall-forming aphids (*Pemphigus*). Evolution 58, 539–553.

Baumann, L., Baumann, P., 2005. Complete mitochondrial genome of *Daktulosphaira vitifoliae*. Available from: <www.ebi.ac.uk>.

Baumann, L., Baumann, P., Moran, N.A., Sandstrom, J., Thao, M.L., 1999. Genetic characterization of plasmids containing genes encoding enzymes of leucine biosynthesis in endosymbionts (*Buchnera*) of aphids. J. Mol. Evol. 48, 77–85.

Brynnel, E.U., Kurland, C.G., Moran, N.A., Andersson, S.G., 1998. Evolutionary rates for tuf genes in endosymbionts of aphids. Mol. Biol. Evol. 15, 574–582.

Clements, K.M., Wiegmann, B.M., Sorenson, C.E., Smith, C.F., Neese, P.A., Roe, R.M., 2000. Genetic variation in the *Myzus persicae* complex (Homoptera: Aphididae): evidence for a single species. Ann. Entomol. Soc. Am. 93, 31–46.

Coeur d'acier, A., Jousselin, E., Martin, J.-F., Rasplus, J.-Y., 2006. Phylogeny of the genus *Aphis* Linnaeus, 1758 (Homoptera: Aphididae) inferred from mitochondrial DNA sequences. Mol. Phylogenet. Evol. 42, 598–611.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Funk, D.J., Helbling, L., Wernegreen, J.J., Moran, N.A., 2000. Intraspecific phylogenetic congruence among multiple symbiont genomes. Proc. Biol. Sci. 267, 2517–2521.

Gao, N., Foster, R.G., Hardie, J., 2000. Two opsin genes from the vetch aphid, *Megoura viciae*. Insect Mol. Biol. 9, 197–202.

Gil, R., Sabater-Munoz, B., Latorre, A., Silva, F.J., Moya, A., 2002. Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. Proc. Natl. Acad. Sci. USA 99, 4454–4458.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Heie, O.E., 1980. The Aphidoidea (Hemiptera) of Fennoscandia and Denmark. I. General Part. The Families Mindaridae, Hormaphididae, Thelaxidae, Anoeciidae, and Pemhigidae. Fauna Entomol. Scand. 9, 1–236.

Heie, O.E., 1987. Palaeontology and phylogeny. In: Minks, A.K., Harrewijn, P. (Eds.), Aphids. Their Biology, Natural Enemies and Control. Elsevier, Amsterdam, pp. 367–391.

Heie, O.E., 2004. The history of the studies on aphid palaeontology and their bearing on the evolutionary history of aphids. In: Simon, J.C., Dedryver, C.A., Rispe, C., Hulle, M. (Eds.), Aphids in a New Millennium. INRA Editions, Paris, pp. 151–158.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Ilharco, F.A., van Harten, A., 1987. Systematics. In: Minks, A.K., Harrewijn, P. (Eds.), Aphids. Their Biology, Natural Enemies and Control. Elsevier, Amsterdam, pp. 51–78.

Inbar, M., Wink, M., Wool, D., 2004. The evolution of host plant manipulation by insects: molecular and ecological evidence from gall-forming aphids on *Pistacia*. Mol. Phylogenet. Evol. 32, 504–511.

Kumar, S., 1995. PHYLTEST: A Program for Testing Phylogenetic Hypothesis. Penn State University, Pennsylvania.

Kumar, S., Tamura, K., Nei, M., 2004. *MEGA3*: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Martínez, D., Moya, A., Latorre, A., Fereres, A., 1992. Mitochondrial DNA variation in *Rhopalosiphum padi* (Homoptera: Aphididae) populations from four Spanish localities. Ann. Entomol. Soc. Am. 85, 241–246.

Martínez-Torres, D., Buades, C., Latorre, A., Moya, A., 2001. Molecular systematics of aphids and their primary endosymbionts. Mol. Phylogenet. Evol. 20, 437–449.

Moran, N.A., Baumann, P., 1994. Phylogenetics of cytoplasmically inherited microorganism of arthropods. Trends Ecol. Evol. 9, 15–20.

Moran, N.A., Munson, M.A., Baumann, P., Ishikawa, H., 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. Lond. B 253, 167–171.

Moran, N.A., Kaplan, M.E., Gelsey, M.J., Murphy, T.G., Scholes, E.A., 1999. Phylogenetics and evolution of the aphid genus *Uroleucon* based on mitochondrial and nuclear DNA sequences. Syst. Entomol. 24, 85–93.

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

Nieto-Nafría, J.M., Mier-Durante, M.P., 1998. Fauna Ibérica. Hemiptera Aphididae I. Museo Nacional de Ciencias Naturales. Consejo Superior de Investigaciones Científicas, Madrid.

Nieto-Nafría, J.M., Mier-Durante, M.P., Remaudière, G., 1997. Les noms des taxa du groupe-famille chez les Aphididae (Hemiptera). Revue fr. Ent. (N.S.) 19, 77–92.

Normark, B.B., 2000. Molecular systematics and evolution of the aphid family Lachnidae. Mol. Phylogenet. Evol. 14, 131–140.

Ortiz-Rivas, B., Moya, A., Martinez-Torres, D., 2004. Molecular systematics of aphids (Homoptera: Aphididae): new insights from the long-wavelength opsin gene. Mol. Phylogenet. Evol. 30, 24–37.

Perez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A., Latorre, A., 2006. A small microbial genome: the end of a long symbiotic relationship? Science 314, 312–313.

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

Quednau, F.W., Remaudière, G., 1994. Le genre sud-américain *Neuquenaphis* E.E. Blanchard, description de deux nouvelles espéces et définition de nouvelles sous-familles ďAphididae (Homoptera). B. Soc. Entomol. Fr. 99, 365–384.

Rambaut, A., Drummond, A., 2003. Tracer. University of Edinburgh.

Remaudière, G., Quednau, F.W., 1988. Description de deux *Pterasthenia* nouveaux et révision de Pterastheniinae subfam. nov. (Hom.: Aphididae). Ann. Soc. Entomol. Fr. 24, 47–57.

Remaudière, G., Remaudière, M., 1997. Catalogue of the World's Aphididae. INRA Editions, Paris.

Remaudière, G., Stroyan, H.L.G., 1984. Un *Tamalia* nouveau de Californie (USA) et discussion sur les Tamalinae subfam. nov. (Hom. Aphididae). Ann. Soc. Entomol. Fr. 20, 93–103.

Robinson-Rechavi, M., Huchon, D., 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics 16, 296–297.

Rouhbakhsh, D., Lai, C.Y., von Dohlen, C.D., Clark, M.A., Baumann, L., Baumann, P., Moran, N.A., Voegtlin, D.J., 1996. The tryptophan biosynthetic pathway of aphid endosymbionts (*Buchnera*): genetics and evolution of plasmid-associated anthranilate synthase (trpEG) within the Aphididae. J. Mol. Evol. 42, 414–421.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502–504.

Shaposhnikov, G.Ch., 1987. Evolution of aphids in relation to evolution of plants. In: Minks, A.K., Harrewijn, P. (Eds.), Aphids. Their Biology, Natural Enemies and Control. Elsevier, Amsterdam, pp. 409–414.

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

Shingleton, A.W., Stern, D.L., 2003. Molecular phylogenetic evidence for multiple gains or losses of ant mutualism within the aphid genus *Chaitophorus*. Mol. Phylogenet. Evol. 26, 26–35.

Silva, F.J., van Ham, R.C., Sabater, B., Latorre, A., 1998. Structure and evolution of the leucine plasmids carried by the endosymbiont (*Buchnera aphidicola*) from aphids of the family Aphididae. FEMS Microbiol. Lett. 168, 43–49.

Staden, R., Beal, K., Bonfield, J.K., 1998. The Staden package. In: Misener, S., Krawetz, S. (Eds.), Computer Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 115–130.

Stern, D.L., 1994. A phylogenetic analysis of soldier evolution in the aphid family Hormaphididae. Proc. R. Soc. Lond. B 256, 203–209.

Stern, D.L., Aoki, S., Kurosu, U., 1997. Determining aphid taxonomic affinities and life cycles with molecular data: a case study of the tribe Cerataphidini (Hormaphididae: Aphidoidea: Hemiptera). Syst. Entomol. 22, 81–96.

Strimmer, K., Rambaut, A., 2002. Inferring confidence sets of possibly misspecified gene trees. Proc. Biol. Sci. 269, 137–142.

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

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

The International Aphid Genomics Consortium, 2009. Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol., in press..

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

van Ham, R.C., Martinez-Torres, D., Moya, A., Latorre, A., 1999. Plasmid-encoded anthranilate synthase (TrpEG) in *Buchnera aphidicola* from aphids of the family Pemphigidae. Appl. Environ. Microbiol. 65, 117–125.

van Ham, R.C., Gonzalez-Candelas, F., Silva, F.J., Sabater, B., Moya, A., Latorre, A., 2000. Postsymbiotic plasmid acquisition and evolution of the repA1-replicon in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. USA 97, 10855–10860.

von Dohlen, C.D., Moran, N.A., 2000. Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. Biol. J. Linn. Soc. 71, 689–717.

von Dohlen, C.D., Teulon, D.A., 2003. Phylogeny and historical biogeography of New Zealand Indigenous Aphidini Aphids (Hemiptera, Aphididae): an hypothesis. Systematics 96, 107–116.

von Dohlen, C.D., Kurosu, U., Aoki, S., 2002. Phylogenetics and evolution of the eastern Asian-eastern North American disjunct aphid tribe, Hormaphidini (Hemiptera: Aphididae). Mol. Phylogenet. Evol. 23, 257–267.

von Dohlen, C.D., Rowe, C.A., Heie, O.E., 2006. A test of morphological hypotheses for tribal and subtribal relationships of Aphidinae (Insecta: Hemiptera: Aphididae) using DNA sequences. Mol. Phylogenet. Evol. 38, 316–329.

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