Master of Science Thesis

Analysis of the crane flies (Diptera, Tipuloidea) subfamilies relationships using molecular sequence data

Saija Ahonen

University of Jyväskylä

Department of Ecological and Environmental Science

Evolution genetic

23.7.2008

JYVÄSKYLÄN YLIOPISTO, Matemaattis-luonnontieteellinen tiedekunta

Bio- ja ympäristötieteiden laitos
Evoluutiogenetiikka

AHONEN, S. : Vaaksiaisten (Diptera, Tipuloidea) alaheimojen sukulaisuussuhteiden selvittäminen molekyylisekvenssiaineiston avulla

Pro Gradu -tutkielma: 28 s.
Työn ohjaajat: Prof. Alessandro Grapputo, Prof. Johanna Mappes
Tarkastajat: Prof. Anneli Hoikkala
Heinäkuu 2008

Hakusanat: Diptera, elongaatio faktori 1α, sukupuu, sytokromi oksidaasi I, sytokromi oksidaasi II, Tipuloidea

TIIVISTELMÄ

Viimeisimpien vuosikymmenien aikana molekylaarisiin aineistoihin perustuvat menetelmät ovat osoittautuneet erittäin hyödyllisiksi organismien sukulaisuussuhteita tutkittaessa. Niiden avulla on mm. paljastettu kryptisiä lajeja, ja ne ovat mahdollistaneet myös hyvin kaukaisten organismien tutkimisen, mikä on ollut vaikeampaa pelkästään morfologisiin piirteisiin perustuen. Vaaksiaiset (Tipuloidea) ovat yksi kaksisiipisten (Diptera) lajirikkaimpia ryhmiä. Siitä huolimatta vaaksiaisista ei ole aiemmin tehty molekylaarisiin menetelmiin perustuvaa alemman tason sukupuututkimusta, eivätkä morfologiaan perustuvat tutkimukset ole antaneet täysin yksiselitteisiä tuloksia vaaksiaisten sukulaisuussuhteista. Tässä tutkimuksessa rakennettiin vaaksiaisten sukupuu käyttäen pätkiä proteiineja koodaavista geeneistä; mitokondriaalisesta sytokromi oksidaasi I (654 bp) ja tumallisesta elongaatio faktori 1α (950 bp) geeneistä. DNA:ta eristettiin kolmeltakymmeneltäviideltä vaaksiaislajilta ja kahdelta ulkoryhmänä toimineilta Trichoceridae -lajilta. Sukupuiden tekoon käytettiin Maximum Likelihood, Maximum Parsimony, Neighbour Joining ja Bayesian analyysimenetelmiä, sekä erillisille että yhdistetylle COI- ja Ef1α-sekvenssiaineistolle. Samalla puiden haarojen luotettavuus arvioitiin. COI- ja Ef1α-aineistojen soveltuvuutta sukupuuanalyyseihin arvioitiin testaamalla sekvenssien kyllästyneisyys. Yhdistetyn COI- ja Ef1α-aineiston perusteella voitiin varmistaa, että vaaksiaisalaheimot Tipulinae ja Cylindrotominae, Uliinae, Pediciinae, Limoniinae ja Chioneinae, ovat monofyleettisiä, kun taas Limoniidae heimon, sekä sen Limnophilinae alaheimon lajit vaikuttaisivat olevan parafyleettisiä. Nämä tulokset vastasivat aiemman vaaksiaisten toukka- ja kotelovaiheiden morfologiaan perustuvan tutkimuksen tuloksia. Tämän tutkimuksen avulla ei kuitenkaan pystytty sen tarkemmin selvittämään sukulaisuussuhteita eri vaaksiaisheimojen tai alaheimojen kesken. Tämä johtui suurelta osalta molemmista aineistoista havaitusta, mutta etenkin COI-sekvenssien kyllästyneisyydestä. Ef1α-aineistolla saatiin sukupuita, joissa puiden haarat olivat paremmin eroteltuina, kun taas COI-aineisto soveltui tässä tutkimuksessa lähinnä vain erittäin läheisten lajien sukulaisuussuhteiden selvittämiseen.

UNIVERSITY OF JYVÄSKYLÄ, Faculty of Science

Department of Ecological and Environmental Science

AHONEN, S.: Analysis of the crane flies (Diptera, Tipuloidea) subfamilies relationships using molecular sequence data

Master of Science Thesis: 28 p.
Supervisors: Prof. Alessandro Grapputo, Prof Johanna Mappes
Inspectors: Prof. Anneli Hoikkala
July 2008

Key Words: cytochrome oxidase I, cytochrome oxidase II, Diptera, elongation factor 1α, phylogeny, Tipuloidea

ABSTRACT

During the last decade, molecular methods have provided a powerful tool for phylogenetic studies. They have revealed cryptic species, and made possible to study quite distantly related organisms, which have been much harder when using only morphological data. Despite crane flies (Tipuloidea) is one of the species richest groups of true flies (Diptera), no molecular based phylogenetic studies have been done on this group so far. Furthermore, the morphological based phylogenetic studies have not given unambiguous results about the relationships among crane flies. Therefore the phylogeny of crane flies was inferred from the fragments of the protein coding mitochondrial cytochrome oxidase I (654 bp) and the nuclear elongation factor 1α (950 bp) genes. Thirty five species from all four crane fly families were sampled; along with two Trichoceridae species used as outgroup. Phylogenetic reconstruction was accomplished by Maximum Likelihood, Maximum Parsimony, Neighbour Joining analyses, and Bayesian inference of separate and combined COI and Ef1α sequence data. The tree topology support was evaluated. The Ef1α and COI sequences were tested for the substitution saturation to define the reliability of the results. The combined data set confirmed the monophyly of the crane fly subfamilies, Tipulinae, Cylindrotominae, Uliinae, Pediciinae, Limoniinae and Chioneinae. While, the monophyly of the family Limoniidae and its subfamily Limnophilinae was not supported by this study. These results support the previous study of crane flies, based on the morphological characters of larvae and pupae. On the contrary, the relationships among crane fly families stayed unresolved or weakly supported. This was mostly due to saturation in both data sets, especially in COI sequences. Thus the Ef1α data gave better resolved trees while the COI sequences were good only in resolving the relationships among very closely related species.

Contents

1. INTRODUCTION

1.1. Molecular data in phylogenetic studies

The purpose of phylogenetic studies are to reconstruct the correct genealogical ties among biological entities, to estimate the time of divergence between organism – i.e., the time since they last shared a common ancestor – and to chronicle the sequence of events along the evolutionary lineages (Graur & Li 2000). These evolutionary relationships are illustrated by means of a phylogenetic tree and taxonomy of organisms is intended to follow it. Before the molecular methods, taxonomy has been based on morphological, anatomical, physiological and paleontological data. However, some organisms could have not been identified correctly by the means of morphological data.

There are several reasons why molecular data is much more suitable for phylogenetic studies than morphological data. Firstly, DNA is strictly heritable entity and the description of molecular characters is unambiguous, whereas some morphological traits can be influenced by environmental factors and descriptions contain ambiguous modifiers as "thin", "reduced" etc. Molecular traits can also provide a clearer picture of the relationships among organisms, since they generally evolve in much more regular manner than morphological characters. Furthermore homology assessment – i.e. if the species or the characters in study have been derived from a common ancestor – is easier with molecular data. There have been developed many sophisticated mathematical and statistical theories that can be used for quantitative analysis of molecular data, while morphological studies concentrate on qualitative features. Some molecular data can be used when studying evolutionary relationships among very distantly related organism as fungi, plants, and animals, and finally, molecular data are much more numerous than morphological data. This makes it very useful when working with organisms such as bacteria, algae and protozoa, which posses only a limited number of morphological or physiological characters that can be used for phylogenetic studies (Graur & Li 2000).

The application of molecular biology techniques and advantages in tree reconstruction methodology have led to tremendous progress in phylogenetic studies, resulting in a better understanding of the evolutionary history of almost every taxonomic group (Graur & Li 2000). Many longstanding issues have been resolved, such as the question about the closest living evolutionary relatives of human. The commonly held view was that humans were phylogenetically distinct from the great apes –chimpanzees (*Pan*), gorillas (*Gorilla*) and orang-utans (*Pongo*) – being placed in different taxonomic families, and that this split occurred at least 15 million years ago. From microcomplement fixation data, Sarich and Wilson (1967) estimated the divergence time between humans and chimpanzees or gorillas to be as recent as 5 million years ago. More recent molecular studies have gone even further by showing that human and chimpanzee are more closely related to each other than either is to the gorilla (Ruvolo 1997).

Another good example, where molecular methods have provided us new information, concerns the traditional view of the living word dividing into two groups – eukaryotes and prokaryotes. It was challenged by Woese and co-workers in their studies based on ribosomal DNA (Woese & Fox 1977, Fox *et al.* 1980). They revealed that there were in fact two very different groups of prokaryotes, the eubacteria and archaebacteria. Furthermore, they found out that eubacteria, archaebacteria and eukaryotes were derived from a common ancestor, and are about equally distant from each other. Thus the present tree of life contains three domains: Eucarya, Bacteria and Archaea, proposed by Woese *et al.* (1990).

Furthermore cryptic species, organisms that appear identical but are genetically quite distinct, have been found by using molecular methods. For example African elephant (*Loxodonta africana*) was thought to be a one species until a molecular based study by Roca *et al.* (2001) revealed that elephants were actually two genetically distinct species, the African bush elephant (*Loxodonta africana*) and the African forest elephant (*L. cyclotis*). Another example of a cryptic species is the neotropical skipper butterfly (*Astraptes fulgerator*), which despite looking identical, turned out to be 10 distinct species after DNA analysis (Hebert *et al.* 2004).

Phylogenetic studies are important also in the case of threatened species and their protection. Taxonomic status must be accurately established so that endangered species are not denied protection, nor effort wasted on abundant species (Frankham *et al.* 2002). Incorrect lumping of several distinct species into one recognised species has denied protection to endangered species. The threatened tuataras in New Zealand were thought to be a single species until the studies with 25 allozyme loci and morphology, revealed that it consisted of two species, *Sphenodon punctatus* and the critically endangered *S. guntheri* (Daugherty *et al.* 1990). Conversely, splitting of one species into two or more recognized taxa may also lead to erroneous conservation decisions. The North American pumas (*Puma concolor*) were traditionally recognized as eight morphological sub-species, including the critically endangered Florida panther (*P. c. coryi*). However, microsatellite and mtDNA analyses found no significant differentiation among the populations (Culver *et al.* 2000). Thus, the Florida panther populations could be saved from extinction by transferring individuals from other North American puma populations.

1.2. Phylogeny of crane flies

Crane flies (Tipuloidea) are a superfamily of long-horned flies (Diptera, Nematocera). The 15,290 species and subspecies (Oosterbroek 2008) make it one of the biggest groups of true flies (Diptera). Crane fly species are found from all over the world, ranging from the equatorial zone to the arctic and from the sea shores to over 5,600 m of high mountain ranges (Alexander & Byers 1981). The Tipuloidea belong to the infraorder Tipulomorpha, which has been considered to be a problematic clade, both in terms of its family composition and its phylogenetic position within the lower Diptera (Yeates *et al.* 2007). Traditionally the Tipuloidea were regarded as one of the most primitive lineages of Diptera (Wood & Borkent 1989). However, other studies support a close relationship with the Psychodomorpha, which are closely related to the higher flies (Brachycera) (Oosterbroek & Courtney 1995). The same authors, among others, have been proposed that the Tipulomorpha contains Tipuloidea and Trichoceridae. On the contrary, later studies have suggested that Tipulomorpha is paraphyletic, i.e. group whose common ancestor is shared by other taxon or taxa (Friedrich & Tautz 1997), and that Trichoceridae nests with Psychodomorpha (Yeates *et al.* 2007). Furthermore even the tipuloidean classification has not reached the full unanimity. Thus while North American authors (Byers 1992) classify the crane flies in only one family, Tipulidae, containing four subfamilies Limoniinae, Tipulinae and Cylindrotominae, Europeans divide the crane flies into 4 families; Limoniidae, Pediciidae, Tipulidae and Cylindrotomidae (de Jong *et al.* 2008). I will use the European classification in this study.

No molecular phylogenetic studies have been conducted so far to determine the relationships among and within the families of Tipuloidea. The monophyly of the crane flies is supported by the analysis of morphological characters of different life stages (Wood & Borkent 1989, Oosterbroek & Theowald 1991, Oosterbroek & Courtney 1995) and the present phylogeny of the crane flies is based on the pupae and larval characters by Oosterbroek & Theowald (1991, Fig. 1a). In this tree the family Pediciidae forms its own

monophyletic clade, subfamily Chioneinae and most of the genera of Limnophilinae form another clade, and the third group contains some Limnophilinae genera, subfamily Limoniinae and families Cylindrotomidae and Tipulidae. However, the crane fly phylogeny by Oosterbroek & Courtney differs from the phylogenetic tree based on the adult morphological characters by Stary (1992, Fig. 1b). In this tree, crane flies are divided into two monophyletic groups, with family Limoniidae forming the sister group of the families Pediciidae, Tipulidae and Cylindrotomidae. Some other morphology based lower level phylogenetic studies of crane flies have mainly concentrated on a certain smaller groups, such as the subgenera of Tipula (de Jong 1994, 1995).

Since there are still some uncertainties in the phylogeny and classification of the crane flies, more phylogenetic research needs to be done and molecular methods can offer a lot of new information for this group of insects.

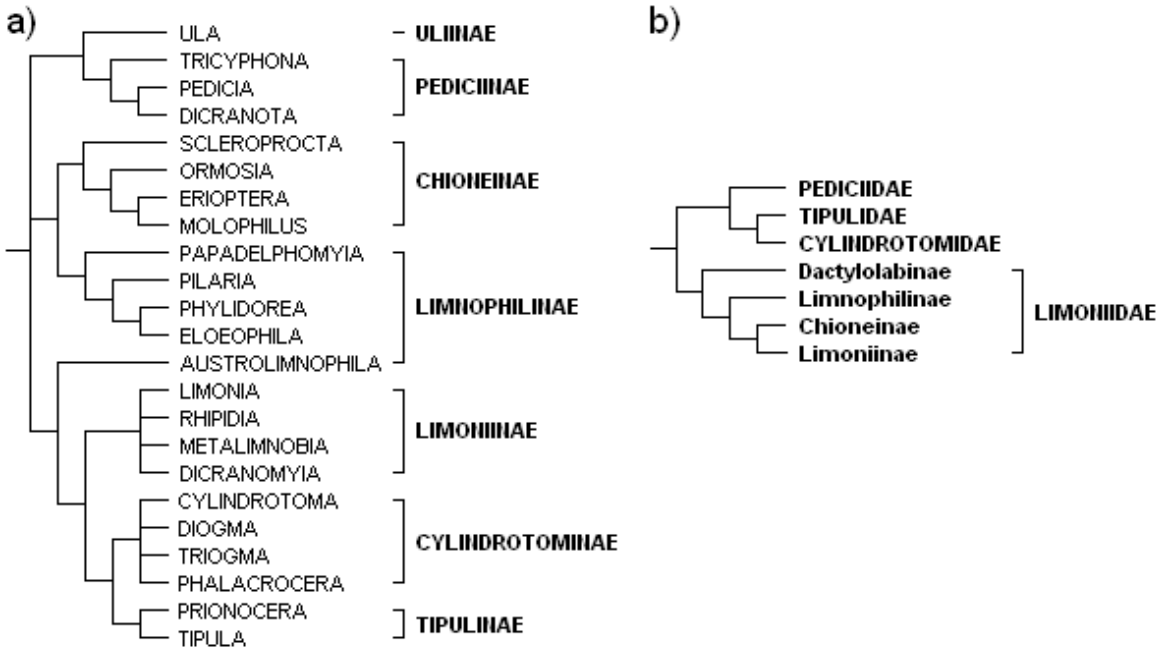

Figure 1. Phylogenetic relationships of Tipuloidea after a) Oosterbroek and Theowald (1991) based on morphological characters of larvae and pupae, (only the genera used in this study are shown in the tree,) and b) Stary (1992) based on adult morphological characters.

1.3. Purpose of the study

The purpose of this study is to infer the phylogenetic relationships among the four crane fly families; Cylindrotomidae, Limoniidae, Pediciidae and Tipulidae, by using DNA sequences of mitochondrial and nuclear genes. Further, the resulting trees obtained from nuclear elongation factor 1α (Ef1α) and the mitochondrial cytochrome oxidase I and II (COI and COII) gene data, will be compared to see if one gives better results. Finally, the trees received from the molecular data are compared with those obtained from morphological data (Oosterbroek & Theowald 1991, Stary 1992) (Fig. 1).

2. MATERIAL AND METHODS

2.1. Study samples

All the 42 crane fly samples (Table 1) were provided by Jukka Salmela, who had also identified the species morphologically. The samples were caught in Finland, most of them

with Malaise trap and the rest with sweep net. The Malaise trap catches were originally preserved in solution of 50% ethylene glycol, water and detergents, and later stored in 70% or 99.5% alcohol. The sweep net catches were preserved immediately in 70% or 99.5 % alcohol. Finally all the samples were stored in 99.5% alcohol. The samples include species from all the families with almost all the subfamilies of Tipuloidea. The Tipulidae subfamilies, Dolichopezinae and Ctenophorinae, and the Limoniidae subfamily Dactylolabinae, were missing from this data. Two species of the family Trichoceridae were selected as outgroup, as this family has been regarded as a sister-group of Tipuloidea in earlier studies (for example Oosterbroek & Courtney 1995).

Table 1. The species used in this study divided into their families and subfamilies.

Cylindrotomidae	
	Cylindrotominae
	Cylindrotoma distinctissima[a] (Meigen, 1818)
	Diogma caudata (Takahashi, 1960)
	Phalacrocera replicata (Linnaeus, 1758)
	Triogma trisulcata (Schummel, 1829)
Limoniidae	
	Chioneinae
	Erioptera (Erioptera) lutea[b] (Meigen, 1804)
	Molophilus (Molophilus) appendiculatus (Staeger, 1840) (two specimens)
	Molophilus (Molophilus) ater (Meigen, 1804)
	Ormosia (Ormosia) depilata (Edwards, 1938) (two specimens)
	Scleroprocta sororcula (Zetterstedt, 1851)
	Limnophilinae
	Austrolimnophila (Archilimnophila) unica (Osten Sacken, 1869)
	Eloeophila maculata (Meigen, 1804) (two specimens)
	Eloeophila trimaculata (Zetterstedt, 1838) (two specimens)
	Paradelphomyia (Oxyrhiza) fuscula (Loew, 1873)
	Phylidorea (Phylidorea) abdominalis (Staeger, 1840)
	Phylidorea (Phylidorea) squalens (Zetterstedt, 1838)
	Pilaria meridiana (Staeger, 1840)
	Limoniinae
	Dicranomyia (Dicranomyia) consimilis (Zetterstedt, 1838)
	Dicranomyia (Dicranomyia) distendens (Lundström, 1912)
	Dicranomyia (Numantia) fusca (Meigen, 1804)
	Limonia flavipes (Fabricius, 1787)
	Metalimnobia (Metalimnobia) bifasciata (Schrank, 1781)
	Rhipidia (Rhipidia) maculata (Meigen, 1818)

Pediciidae
Pediciinae
Dicranota (Dicranota) guerini[b] (Zetterstedt, 1838)
Dicranota (Paradicranota) pavida (Haliday, 1833)
Pedicia (Crunobia) straminea (Meigen, 1838)
Pedicia (Pedicia) rivosa (Linnaeus, 1758)
Tricyphona (Tricyphona) immaculata (Meigen, 1804)
Tricyphona (Tricyphona) unicolor (Schummel, 1829)
Uliinae
Ula (Ula) mixta (Stary, 1983) (two specimens)
Ula (Ula) sylvatica (Meigen, 1818) (two specimens)
Tipulidae
Tipulinae
Prionocera turcica (Fabricius, 1787)
Tipula (Savtshenkia) grisescens (Zetterstedt, 1851)
Tipula (Savitshenkia) subnodicornis (Zetterstedt, 1851)
Tipula (Vestiplex) nubeculosa (Meigen, 1804)
Tipula (Vestiplex) scripta (Meigen, 1830)
Trichoceridae (two specimens) not identified as species level

[a] Species only in COI sequence data.
[b] Species only in EF1α sequence data.

2.2. DNA extraction, PCR amplification and sequencing

Extraction of the total genomic DNA was performed following the DNeasy Tissue Kit (Qiagen) protocol for isolation of genomic DNA from insects modified for the KingFisher apparatus (Thermo scientific).

For each sample, fractions of mitochondrial DNA (mtDNA) – including parts of the cytochrome oxidase I (COI) and II (COII) genes – were selected for amplification. The first mtDNA fragment (1,007 bp; amplified with the primers m13r-LCOI1490 and HCOI2198) contained part from the 5' end of the COI gene. The second fragment (708 bp; amplified with the COI2382 and m13f-A3389 primers) and third fragment (598 bp; amplified with the S2792 and m13f-A3389 primers) included part from the 3' end of the COI to the 5' end of the COII gene. Furthermore, a portion (1,067 bp) of the nuclear gene for the elongation factor 1α (Ef1α), was amplified with polymerase chain reaction (PCR). The primers used for the PCR amplification and sequencing are listed in Table 2.

Mitochondrial and nuclear genes have both advantages and disadvantages in phylogenetic studies. MtDNA is usually easier to amplify and conserved mitochondrial primers are widely available (see Simon *et al.* 1994). MtDNA does not have introns, has no recombination, and evolves at high rates. However, high rate of substitution can be disadvantageous when we aim to resolve divergences of more than 5–10 million years. At this level of divergence mtDNA may present high levels of homoplasy, because of substitutional biases. Nuclear genes generally evolve more slowly than mitochondrial genes and usually show less biased base composition, making them better markers for deep

divergences (Lin & Danforth 2004). The Ef1α gene is also highly conserved and free of internal repeats (*Cho et al.* 1995). We chose these genes because of their proven utility in several phylogenetic studies within especially butterflies (Lepidoptera) (for example Monteiro & Pierce 2001, Kandul *et al.* 2004, Silva-Brandão *et al.* 2005).

Table 2. Primers for amplification[1] and sequencing[2] of mitochondrial DNA (COI) and nuclear DNA (EF1α).

Gene	Name	F/R[a]	Location[b,c]	Sequence (5' → 3')
COI	m13r-LCOI1490[1]	F	1490	m13r-GGTCAACAAATCATAAAGATATTGG
	HCOI2198[1,2]	R	2198	TAAACTTCACGGTGACCAAAAAATCA
	COI2382[1,2]	F	2382	ATGCTATACTAGCTATTGG
	S2792[1,2]	F	2792	ATACCTCGACGTTATTCAGA
COII	m13f-A3389[1]	R	3389	TCATAAGTTCARTATCATTG
EF1α	m13r-M44.1[1]	F	2277	m13r-GCTGAGCGYGARCGTGGTATCAC
	rcM4[1,2]	R	3344	ACAGCVACKGTYTGYCTCATRTC
	m13r[2]	F		TTTCACACAGGAAACAGCTATGAC
	m13f[2]	R		CGACGTTGTAAAACGACGGCCATG

Y = C/T; R = G/A; K = G/T; V = C/G/A

[a] F = Forward; R = Reverse

[b] Position from 3'end relative to *Drosophila yakuba* (Clary & Wolstenholme 1985) for COI and COII primers. Primers obtained from Folmer *et al.* (1994) and Simon *et al.* (1994).

[c] Position from 3'end relative to *Drosophila melanogaster* for EF1α primers. Primers obtained from Cho *et al.* (1995).

Polymerase chain reaction (PCR) for amplification of mitochondrial and nuclear genes, was done in 25µl reaction volume containing 20-50ng of DNA, 1 U of Taq DNA polymerase (Biotools), 2.5mM of $MgCl_2$, 1 x PCR Buffer, 5 pmole of each primer, 2.5µM of dNTP. For the Ef1α I used a touch down PCR. The amplification consisted of a denaturation step at the 94ºC for 3 min followed by 10 cycles of 45 s at 94ºC, 30 s annealing temperature (starting at 66ºC minus 1.1ºC at each cycle) and 72ºC for 2 min, followed by 36 cycles of 45 s at 94 ºC, 30 s at 55ºC and 2 min at 72 ºC for. A final elongation step was done at 72ºC for 5 minutes. For the mitochondrial DNA the PCR amplification consisted of a denaturation step at 94ºC for 4 min followed by 36 cycles of 30 s at 94ºC, 30 s at 45ºC and 2 min at 72ºC. A final elongation step was done at 72ºC for 5 minutes.

A fraction of the products of amplification were run on a 1% agarose gel, stained with ethidiumbromide or syber safe, and visualised under UV-light to confirm the amplification success. Amplification products were purified with ExoSap-it and sequencing reactions were done using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystem). Sequencing reaction products were purified with ethanol - ethylene diamine tetraacetic acid - sodium acetate (EtOH-EDTA-NaAC) precipitation and run on a Prism® 3100xl Genetic Analyser sequencer (Applied BioSystems). All PCR amplicons were sequenced multiple times from both directions to ensure high quality reads.

The sequences were edited and aligned using the CodonCode Aligner software (CodonCode Corp.). Ef1α gene occurs as two copies in some Diptera, such as in Drosophila and in honeybees (*Apis mellifera*) (Danforth & Ji 1998). For this reason, if the Ef1α sequences showed any intraspecific variation and heterozygosity at any position, I

created all possible haplotypes to check that they were forming a group in the trees. This would prove that the variation was just a variation at the gene and not due to duplication.

2.3. Phylogenetic analyses

I partitioned the data accordingly to the two amplification fragments: the mtDNA fragment and the nuclear DNA fragment. Only one of the mtDNA fragments – the one containing the part from the 5' end of the COI gene – was used for the phylogenetic analyses. Furthermore, I made the combined data set, by placing the Ef1α sequences after the COI sequences. For the separated analyses of COI and Ef1α sequences, I used only the samples that I was able to sequence, whereas for the combined data set, whenever there was COI or Ef1α sequence missing from one species, I replaced that sequence with equal length of N-characters. If there were more than one haplotypes in the Ef1α sequence, I chose one of them to make the combined data for the species.

The reliability of results from molecular phylogenetics of sequence data depends whether some or all sequences in the data set have already lost phylogenetic information due to substitution saturation. I analyzed saturation of variation both by plotting number of transitions or transversions against F84 distance, and using an entropy-based index of substitution saturation (an index of substitution saturation and its application Xia *et al.* 2003) as implemented in the DAMBE v.5.0.16 program package (Xia & Xie 2001).

The Modeltest 3.7 (Posada & Crandall 1998) and MrModeltest 2.2 (Nylander 2004) programs, with MrMTgui v1.0 interface (Nuin 2008), were used to ascertain the best fit model of nucleotide substitution for nuclear and mitochondrial regions separately and combined.

The phylogenetic analyses were performed with PAUP*4.0b.10 (Swofford 2003), using Maximum Parsimony (MP) and Neighbour Joining (NJ) methods. Bayesian inference (BI) analysis was carried out with MrBayes v3.1.2 (Huelsenbeck & Roquist 2001). Maximum Likelihood (ML) analysis was done using the PhyML online web server (Guidon *et al.* 2005). All the analyses were performed for both genes separately and for combined COI and Ef1α data set. The GTR+i+g was used as the best fitting substitution model for the ML, NJ and BI analyses. Trichoceridae was used as outgroup to root the trees.

Maximum parsimony analyses (MP) were performed using heuristic search with 10 random addition sequence replicates, TBR (tree-bisection-reconnection) branch-swapping and all characters equally weighted. A strict consensus tree was computed when multiple equally parsimonious trees were obtained. The robustness of each branch was determined using the bootstrapping (Felsenstein 1985) with 500 replicates. The Neighbour Joining (NJ) analyses were carried out using heuristic search and the robustness of the NJ trees were tested with 500 bootstrap replicates. The Maximum Likelihood (ML) analyses were performed using the simple hill-climbing algorithm and the robustness of the ML trees was estimated with 100 bootstrap replicates.

When discussing the results I will refer to the bootstrap values as giving weak, moderate, good and strong support. In this study I define "weak support" as bootstrap values of 50-61%, "moderate support" as values 62-74%, "good support" as values 75-88% and "strong support" as values 89-100% following Silva-Brandão *et al.* (2005).

Bayesian analyses were done with 2×10^6 generations for the COI gene and the combined data set, and 1.5×10^6 generation for the Ef1α gene, and sampling trees every 100 cycles. The first 1000 trees were discarded as "burn in" for COI and Ef1α genes, and the first 5000 for the combined data set.

3. RESULTS

3.1. Sequence analyses

The 5' end of the (position from 1490 to 2198) COI gene was sequenced successfully for thirty six species including the two outgroup species. The Ef1α gene was instead sequenced successfully for thirty seven species. The mtDNA fragments (position from 2382 to 3389 and from 2792 to 3389) from the 3' end of the COI to the 5' end of the COII gene were successfully sequenced only for nineteen species (twelve sequences of 2382-3389 and seven sequences of 2792-3389), and therefore they were excluded from the final phylogenetic analyses. After aligning and editing the COI sequences consisted a fragment of 654 bp, and the Ef1α sequences of 950 bp. Combined data set had a total of 1604 bp. A single-codon gap was found in *Pilaria meridiana* COI sequence at the position 351-353 in relation to *Drosophila yakuba* sequence. Alignment of Ef1α did not show any indels. One Ef1α sequences from *Cylindrotoma distinctissima*, and two COI sequences from *Erioptera lutea* and *Dicranota guerini* did not succeed. These missing sequences were replaced with N-characters in the combined data set.

The COI sequences had a total of 285 variable sites and 250 parsimony informative sites. The Ef1α had 393 variable sites and 375 parsimony informative sites. The combined data set had a total of 678 variable characters and 582 parsimony informative characters.

Eighteen of Ef1α sequences showed heterozygote positions. Given the possibility that this locus is duplicated, as has been found in other insects, I artificially constructed all the possible haplotypes from the heterozygote position to see if I had sequenced two different loci or just two possible alleles of the same locus. All the artificial haplotypes from the same individual clustered together in the phylogenetic trees. This could mean that duplication events have not occurred in the Tipuloidea, or the primers I used are specific only for one locus. Alternatively the duplications are so young and interesting single phylogenetic lineages that they do not interfere with the relationships among the species.

Table 3. Two tailed t-test measuring if the observed index of substitution saturation (I_{ss}) is significantly lower than critical I_{ss} value ($I_{ss.c}$), assuming a symmetrical topology and an unsymmetrical topology with different numbers of operational taxonomic units (N_{OTU}) for COI and Ef1α genes.

			Symmetrical topology			Unsymmetrical topology		
Gene	N_{OTU}	I_{ss}	$I_{ss.c}$	T^a	P	$I_{ss.c}$	T^a	P
COI	4	0.201	0.804	34.503	0.000	0.770	32.536	0.000
	8	0.201	0.764	29.424	0.000	0.654	23.668	0.000
	16	0.203	0.737	26.518	0.000	0.533	16.390	0.000
	32	0.206	0.718	23.904	0.000	0.402	9.542	0.000
Ef1α	4	0.167	0.817	48.387	0.000	0.784	45.893	0.000
	8	0.160	0.784	43.257	0.000	0.678	35.896	0.000
	16	0.162	0.762	40.030	0.000	0.567	26.985	0.000
	32	0.164	0.744	38.014	0.000	0.442	18.194	0.000

[a] $df_{COI} = 653$ and $df_{EF1\alpha} = 949$

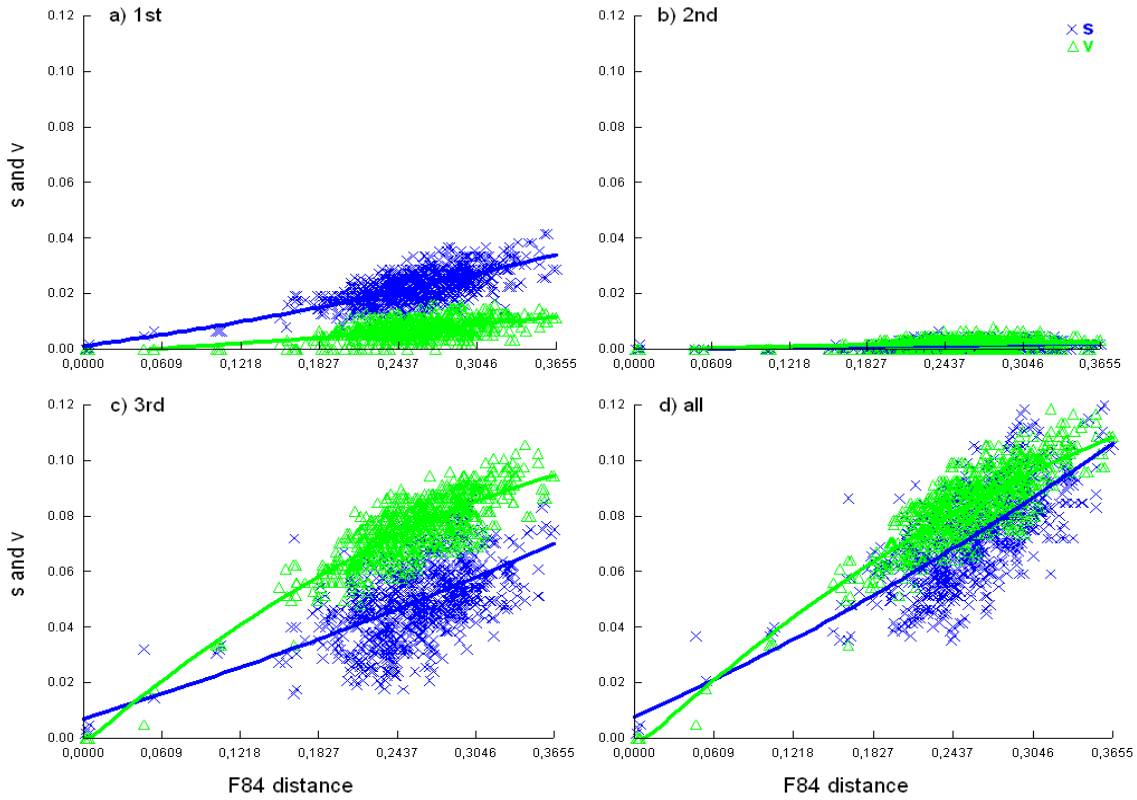

Figure 2. Transition (s, blue) and transversion (v, green) saturation plot for first (a), second (b), third (c) and all codon positions (d) of COI sequences.

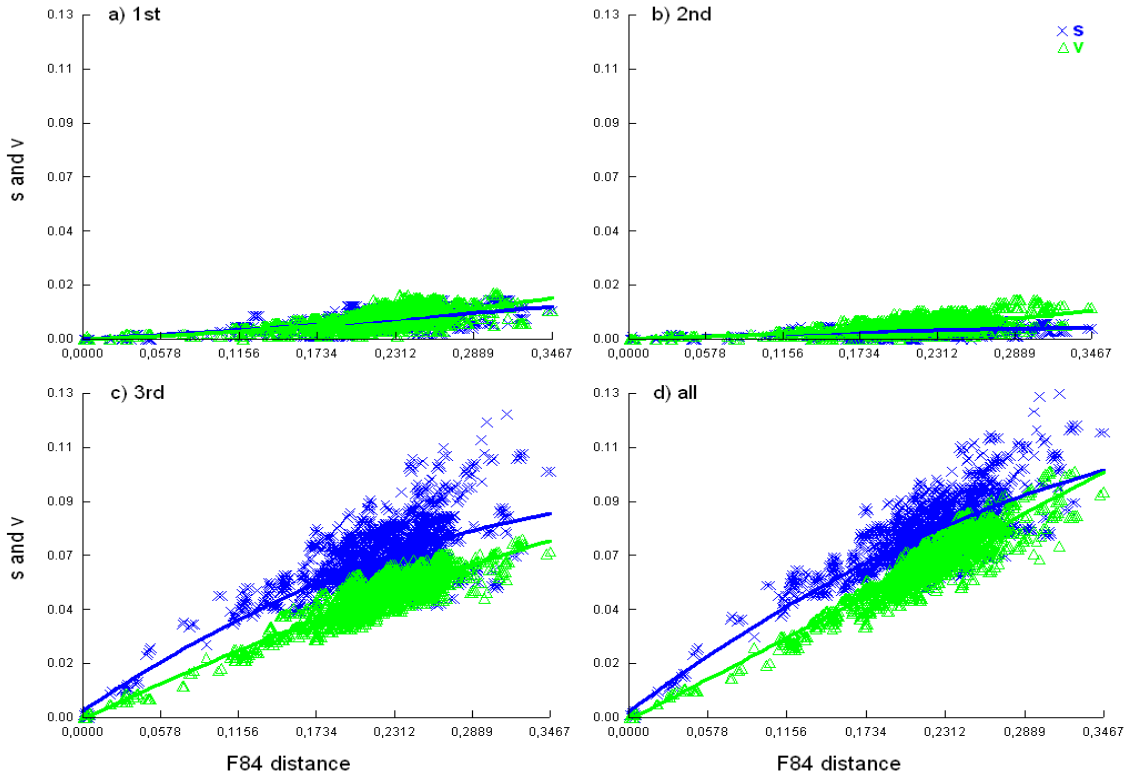

Figure 3. Transition (s, blue) and transversion (v, green) saturation plot for first (a), second (b), third (c) and all codon positions (d) of EF1α sequences.

The substitution saturation was measured by testing whether the observed index of substitution saturation (I_ss) is significantly lower than critical I_{ss} value ($I_{ss.c}$) – at which the sequences will begin to fail to recover the true tree – meaning that if it was significantly lower, then there is a little saturation in the sequences This was done assuming a symmetrical topology and an unsymmetrical topology with different numbers of operational taxonomic units (N_{OTU}) and the results can be seen in Table 3. For the COI sequences, the I_{ss} value is 0.206 in the case of N_{OTU} being 32. The critical $I_{ss.c}$ value is 0.718 if the true tree is symmetrical, and 0.402 if the true tree is asymmetrical, both being highly significantly greater ($p = 0.000$, two tailed t-test) than the observed I_{ss} values. For Ef1α sequences, the $I_{ss.c}$ is 0.744 when assuming a symmetrical topology, and 0.442 when assuming an unsymmetrical topology, were also highly significantly greater ($p = 0.000$) than the observed I_{ss} value 0.164, when N_{OTU} is 32. Thus, there is little substitution saturation at the COI and Ef1α sequences according to the test by Xia.

The other test – the transition (Ts) and transversions (Tv) saturation plot graphics (Figs. 2 and 3) – showed a lot of saturation in the third codon positions of COI and some saturation also in the third codon positions of Ef1α sequences. Figure 2d shows how Ts substitutions accumulate almost linearly with increasing genetic distance between pairs of COI sequences, whereas the Tv substitutions do not, slowly beginning to reach the saturation plateau. This apparent slow-down in the accumulation of substitutions, when increasingly distant sequences are compared, is due to the correspondingly greater saturation level of the more variable third codon positions in the protein-coding sequences analyzed in this study. The Ef1α sequences show reversed pattern in the Figure 3d, where Ts substitutions seems to be reaching the saturation plateau before the Tv substitutions. There are also little substitutions taking place in the first codon positions (Figs. 2a) in the COI sequences, but overall the substitution rates are highest in third position sites (Figs. 2c and 3c), and lowest in second position sites (Figs. 2b and 3b) in both data sets.

3.2. Phylogenetic analyses

The best-fit model for all data sets was GTR+i+g selected by AIC (Akaike information criteria) in Modeltest 3.7 and MrModeltest 2.2 for the combined data, and by hLRT (hierarchical likelihood ratio tests) in Modeltest 3.7 and by AIC in MrModeltest 2.2 for the COI and Ef1α data sets. For all the data sets the number of substitution types was 6. For COI the substitution rate matrix was (1.7105, 8.0672, 5.0870, 1.9025 and 57.2269), for Ef1α (3.2757, 8.7833, 6.1722, 2.3261 and 19.8926) and for combined data set (2.4186, 9.0882, 7.6923, 2.4017 and 22.3558). Assumed nucleotide frequencies and among site rate variation; proportion of invariable sites and distribution of rates at variable sites with shape parameter, are shown in the Table 4 for each data partition.

Table 4. Model parameters for each data partition (Gene): best nucleotide substitution model (Best model), assumed nucleotide frequencies (Base frequency), proportion of invariable sites (i), and shape of the gamma distribution (Gamma).

Gene	Best model	Base frequency				i	Gamma
		A	C	G	T		
COI	GTR+i+g	0.3537	0.1041	0.1227	0.4195	0.5094	0.5745
Ef1α	GTR+i+g	0.3209	0.2086	0.1843	0.2862	0.5234	1.0532
COI+Ef1α	GTR+i+g	0.3158	0.1779	0.1656	0.3407	0.5297	0.9946

GTR+i+g: General Time-Reversible model (Rodríques *et al.* 1990), with proportion of invariable sites (i), and with gamma distribution (Γ).

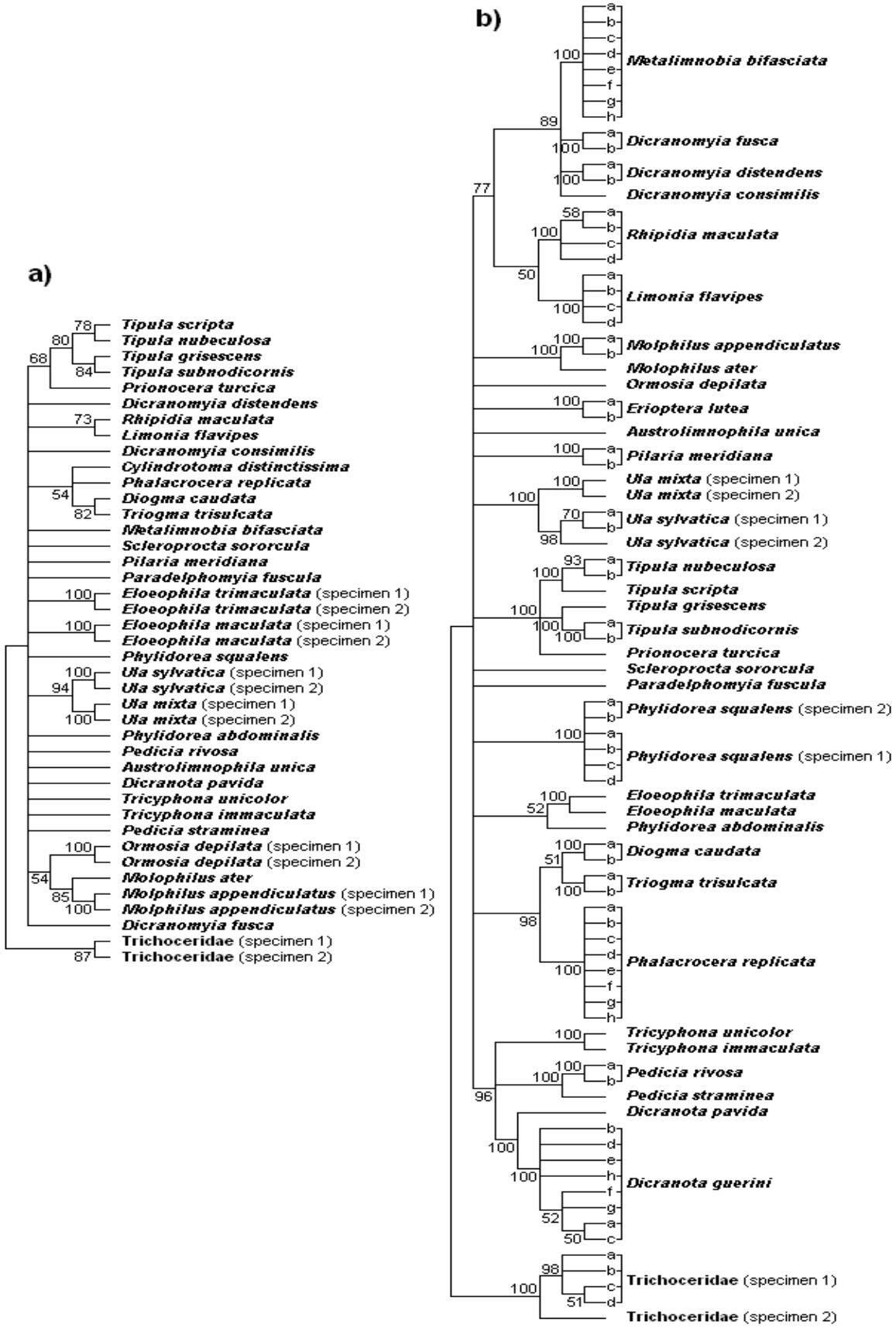

Figure 4. Strict consensus tree obtained from Maximum Likelihood analysis using a) COI and b) EF1α nucleotide sequence. Values under and above the branches indicate the bootstrap support obtained with 100 replicates. (Letters from a to h indicates the artificially constructed Ef1α haplotypes.)

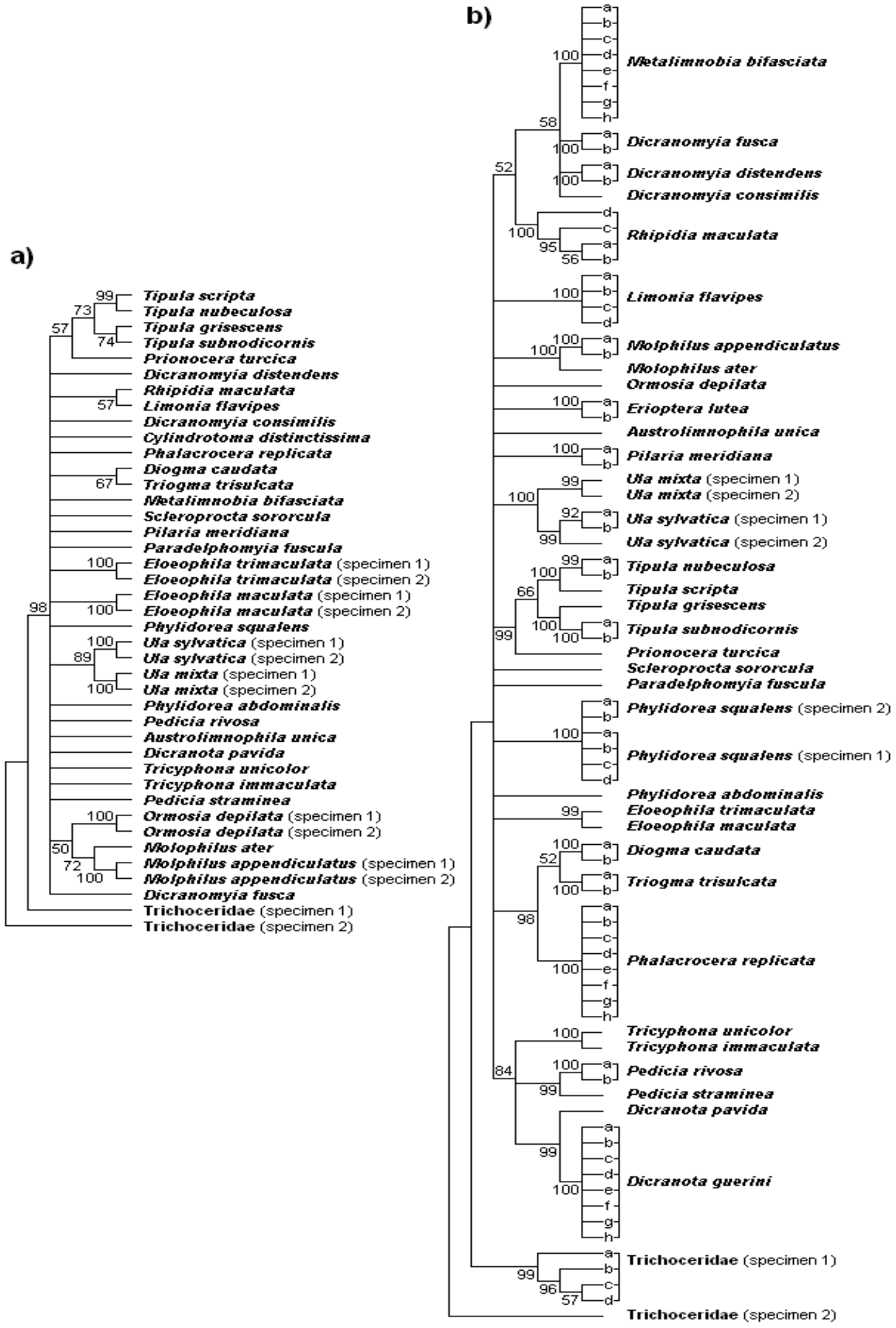

Figure 5. Strict consensus tree obtained from Maximum Parsimony analysis using a) COI and b) EF1α nucleotide sequence. Values under and above the branches indicate the bootstrap support obtained with 100 replicates. (Letters from a to h indicates the artificially constructed Ef1α haplotypes.)

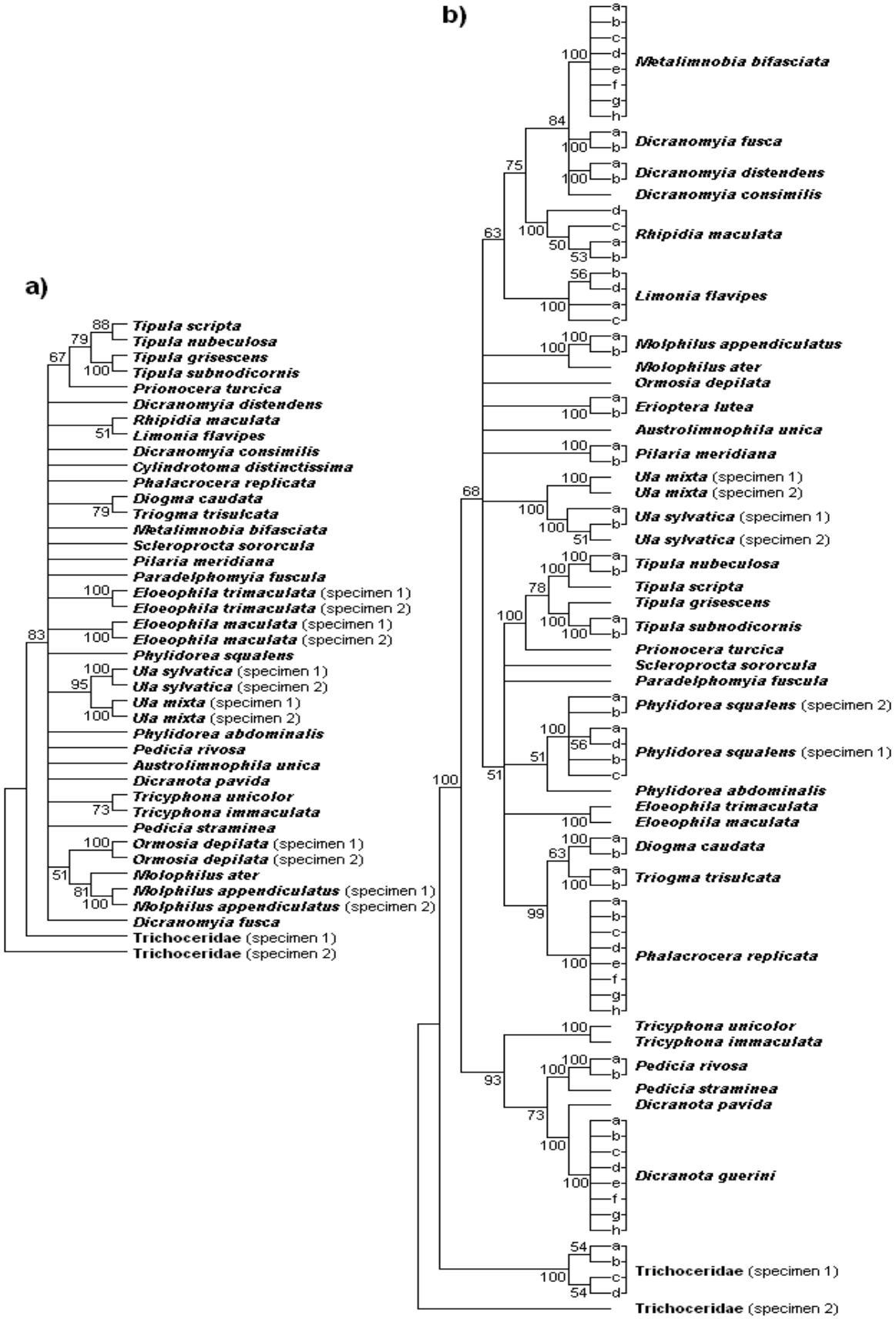

Figure 6. Strict consensus tree obtained from Neighbour Joining method using a) COI and b) EF1α nucleotide sequences. Values under and above the branches indicate the bootstrap support obtained with 100 replicates. (Letters from a to h indicates the artificially constructed Ef1α haplotypes.)

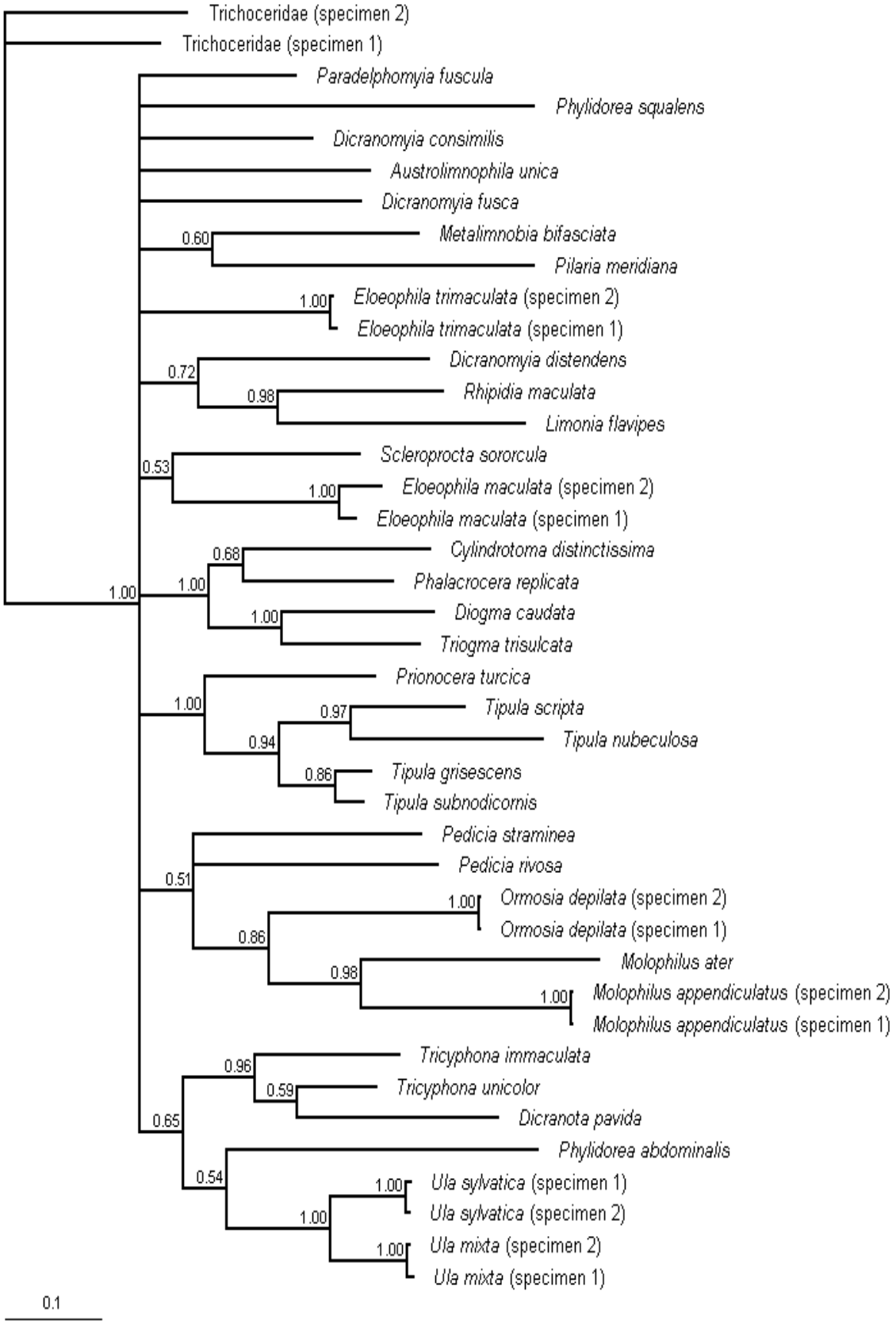

Figure 7. Strict consensus tree inferred by Bayesian analysis from COI nucleotide sequence. Values above the branches indicate Bayesian posterior probability. Scale bar represents expected number of substitutions per site.

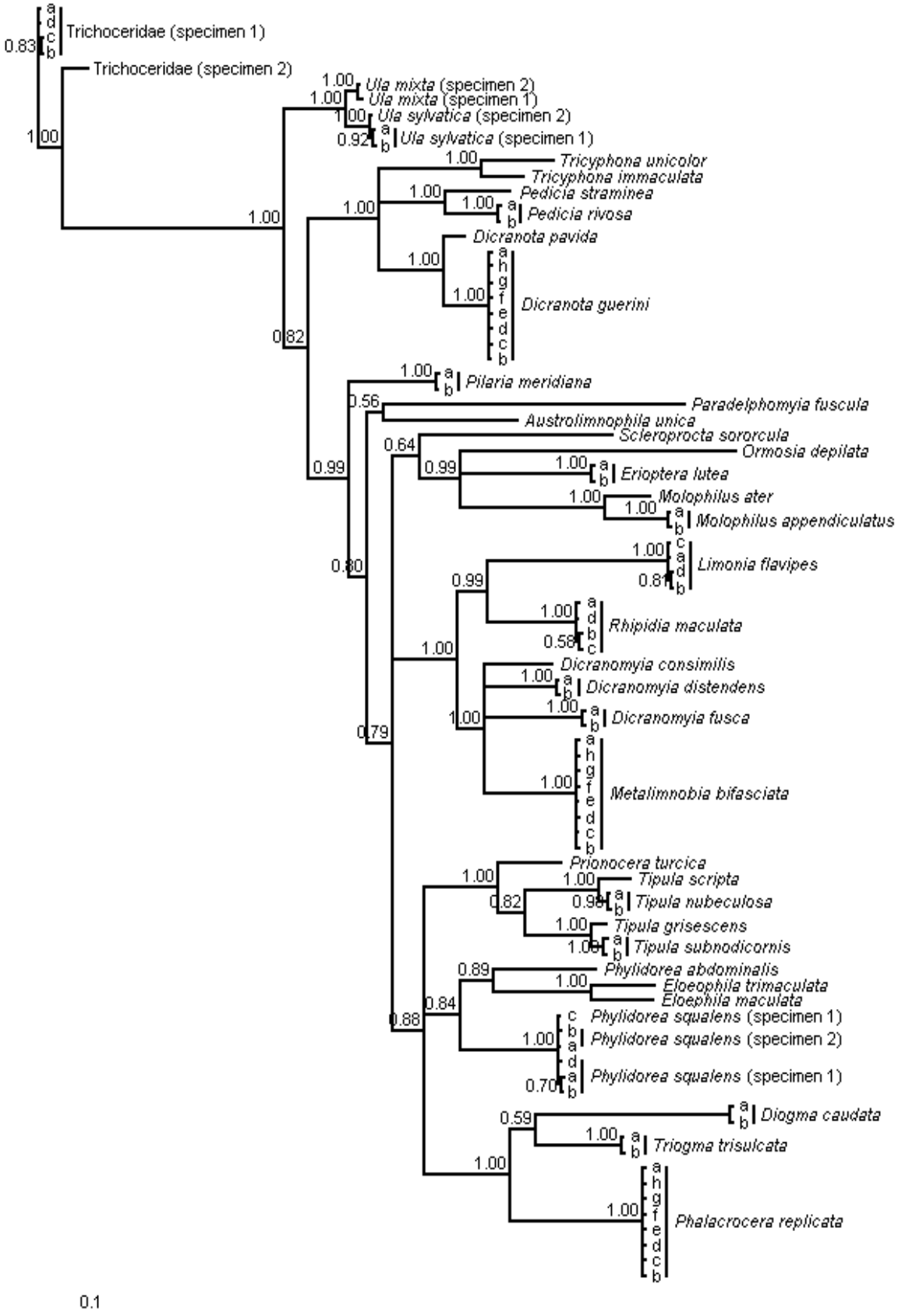

Figure 8. Strict consensus tree inferred by Bayesian analysis from Ef1α nucleotide sequence. Values above the branches indicate Bayesian posterior probability. Scale bar represents expected number of substitutions per site. (Letters from a to h indicates the artificially constructed Ef1α haplotypes.)

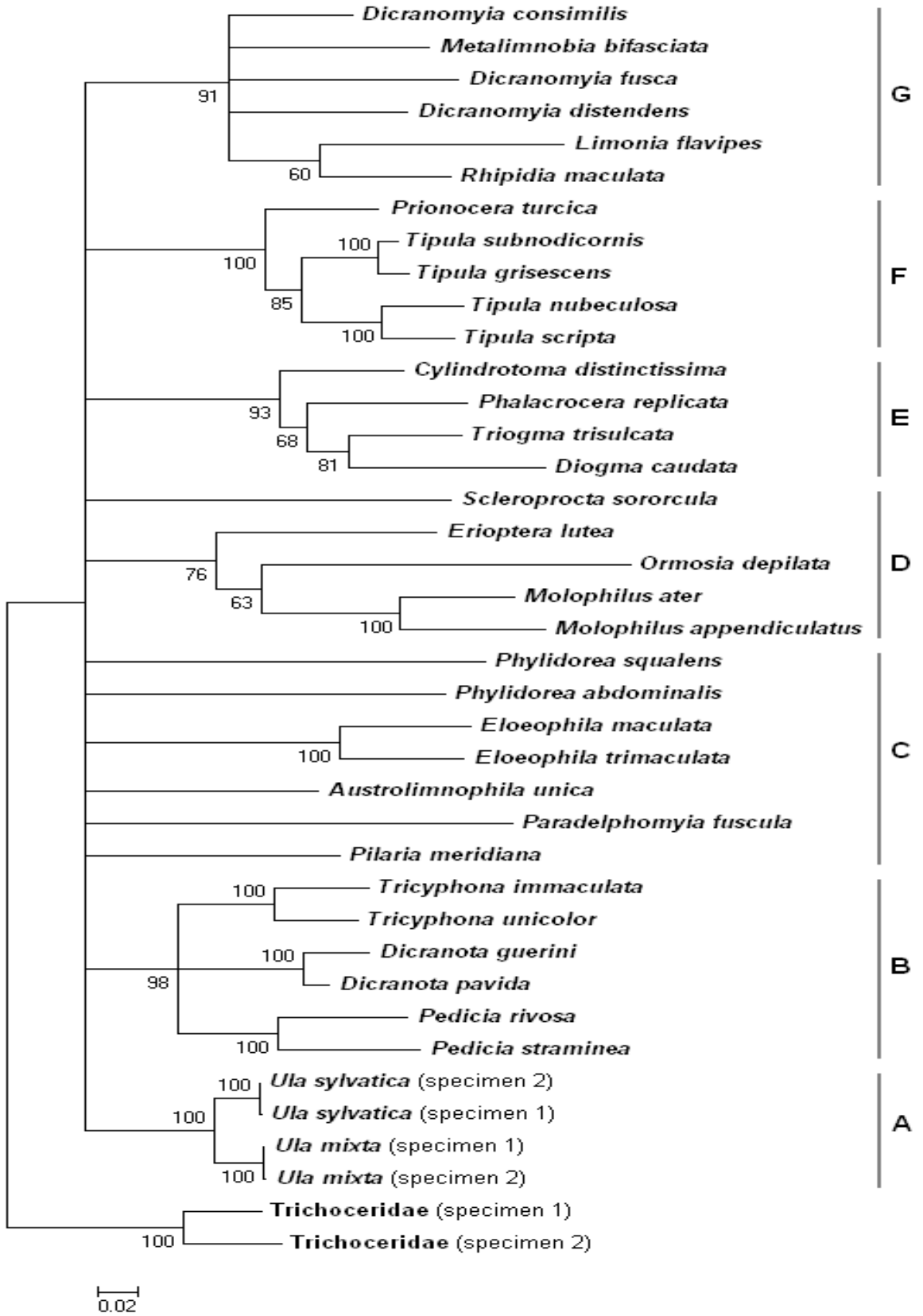

Figure 9. Strict consensus tree obtained from Maximum Likelihood method using combined data. Subfamilies of Tipuloidea are indicated as A) Uliinae, B) Pediciinae, C) Limnophilinae, D) Chioneinae, E) Cylindrotominae, F) Tipulinae and G) Limoniinae. Values under and above the branches denote the bootstrap support. Scale bar represents expected number of substitutions per site.

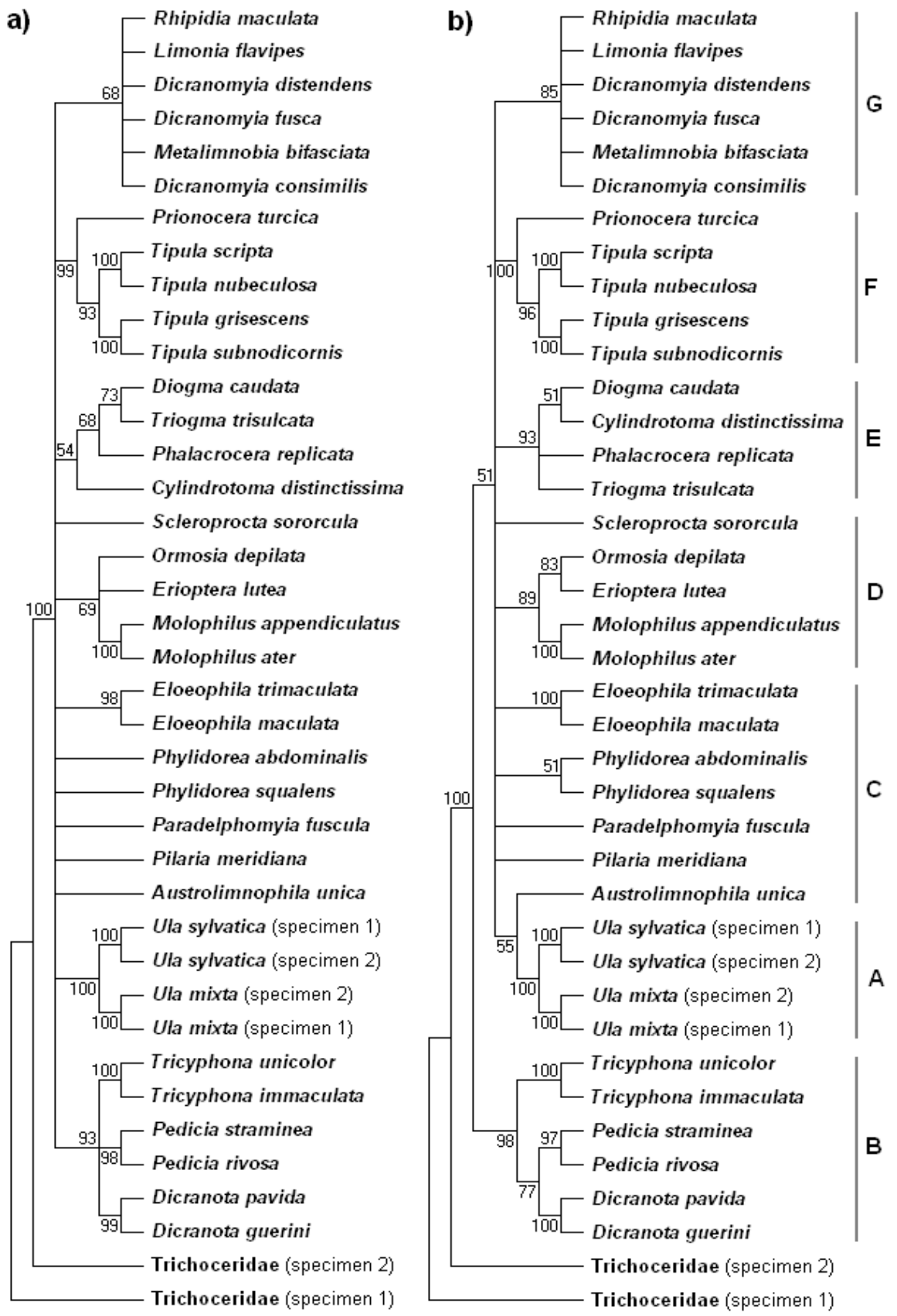

Figure 10. Strict consensus tree obtained from a) Maximum parsimony and b) Neighbour Joining methods using combined data. Subfamilies of Tipuloidea are indicated as A) Uliinae, B) Pediciinae, C) Limnophilinae, D) Chioneinae, E) Cylindrotominae, F) Tipulinae and G) Limoniinae. Values under and above the branches indicate the bootstrap support.

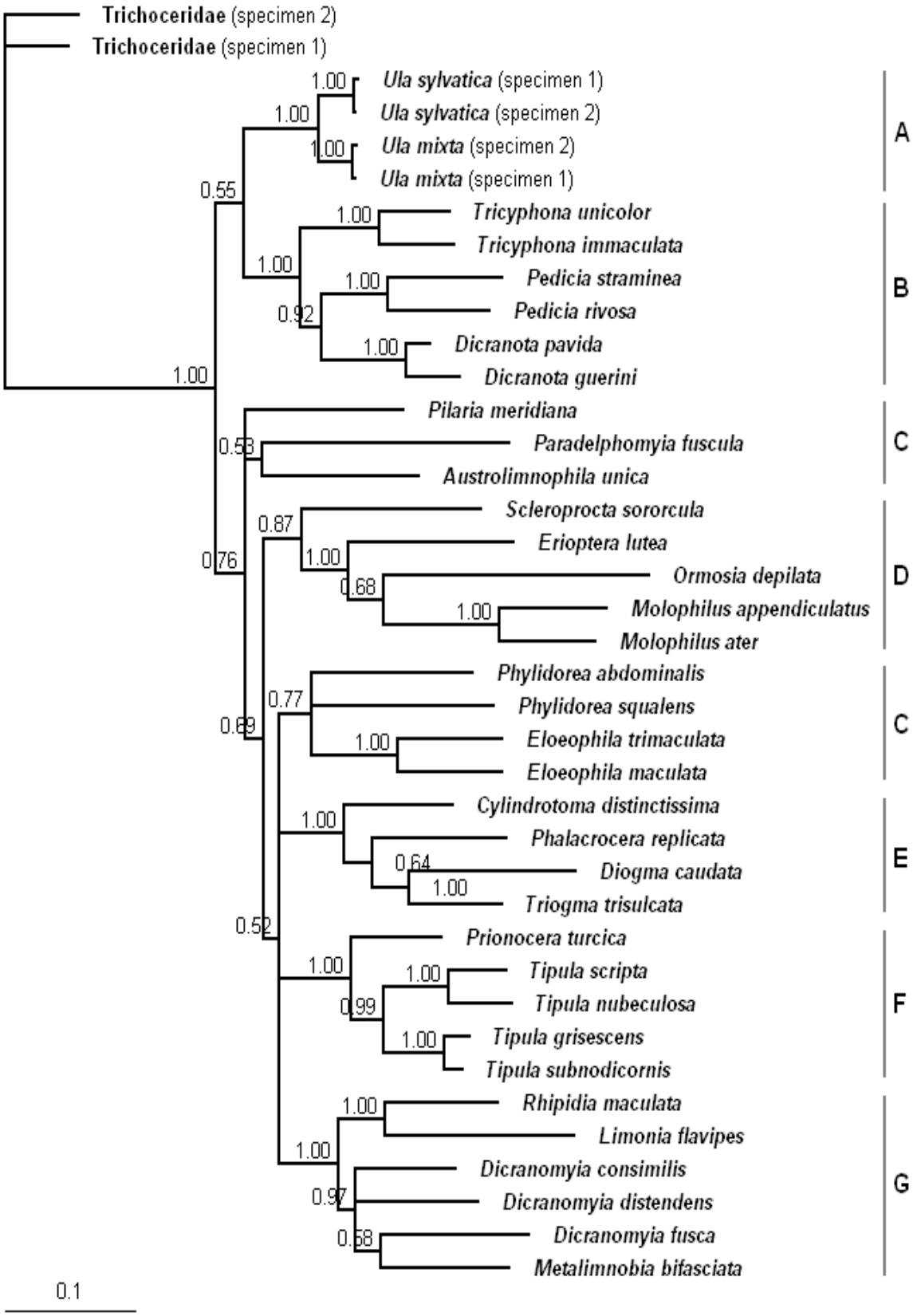

Figure 11. Strict consensus tree inferred by Bayesian analysis from combined data. Subfamilies of Tipuloidea are indicated as A) Uliinae, B) Pediciinae, C) Limnophilinae, D) Chioneinae, E) Cylindrotominae, F) Tipulinae and G) Limoniinae. Values above the branches denote Bayesian posterior probability. Scale bar represents expected number of substitutions per site.

The trees obtained from COI and Ef1α sequence data were quite different (Figs. 4 to 8). In the COI trees almost all the basal nodes were collapsed, only the very closely related species where nicely branched together, giving very little information about the species relationships. Ef1α sequence data gave more resolved trees with all methods and even the tip nodes had equal or better bootstrap of Bayesian posterior probability (pP) support comparing to the COI trees. The COI strict majority rule consensus trees obtained from Maximum likelihood (ML) (Fig. 4a) and Maximum parsimony (MP) (Fig. 5a) were identical, except for the bootstrap values. Comparing to MP and ML trees, Neighbour joining (NJ) tree (Fig. 6a) had only one difference. In NJ tree *Tricyphona immaculata* and *Tricyphona unicolor* were grouped together with moderate bootstrap support. The EF1α trees, obtained with ML (Fig. 4b), MP (Fig. 5b) and NJ (Fig. 6b) methods, had minor differences between them. Both Bayesian (BI) majority rule trees, COI (Fig. 7) and EF1α (Fig. 8), were more resolved than MP, ML and NJ trees from the same gene.

The combined data set gave well resolved trees with all methods (Figs. 9 to 11); still some nodes in the trees were collapsed and could not give reliable information of the phylogenetic relationships among the species. The Bayesian analysis gave a better resolved tree (Fig. 11) than other methods: ML (Fig. 9), MP (Fig. 10a) and NJ (Fig. 10b). Nevertheless the Bayesian posterior probabilities were not high for all nodes. The trees obtained from the combined data with the different methods were very similar, with minor differences in support values.

According to the combined data, the crane fly subfamilies Pediciinae and Uliinae are both monophyletic, and their relationship is supported by strong bootstrap and pP values. Within the Pediciinae the three different genera I analysed form three monophyletic groups (with strong bootstrap and pP values). The genus *Pedicia* is the sister group of *Dicranota* (pP = 0.98 and NJ bootstrap value = 77) while more basal group in the subfamily resulted *Tricyphona*. In the Bayesian tree, the Pediciinae and Uliinae subfamilies are grouped together, but the pP is quite low (0.55) to make reliable conclusions on the family Pediciidae being monophyletic.

The subfamilies Tipulinae and Cylindrotominae are both monophyletic in all combined data trees obtained with different methods. The Tipulidae genus *Prionocera* is sister group of the genus *Tipula*, which is divided in two groups (with strong bootstrap and pP support). The first *Tipula* group includes *Tipula scripta* and *T. nubeculosa*, which belong to the subgenus *Vestiplex*. The second group contains *T. grisescens* and *T. subnodicornis*, which belong to the subgenus *Savtshenkia*. According to ML and Bayesian analyses (with moderate to strong bootstrap and pP support), the Cylindrotomidae genera *Diogma* and *Triogma* form a clade, genus *Phalacrocera* being their sister group, and genus *Cylindrotoma* basal to all of them. MP analysis had weaker bootstrap support and NJ analyses did not give reliable information on these genera. Thus there cannot be made more precise conclusions of the relationships among the Cylindrotomidae genera with this study.

The family Limoniidae is paraphyletic, as its subfamilies – Chioneinae, Limoniinae and Limnophilinae – and their genera being placed in different parts of the tree according to the Bayesian analysis for combined data. The subfamily Chioneinae forms a monophyletic group only in Bayesian combined data analysis. One of the Chioneinae genera *Molophilus* forms a monophyletic clade with very strong (100% bootstrap and pP) support in all trees. Genera *Ormosia* and *Erioptera* could be regarded as sister groups of *Molophilus* (with moderate to good bootstrap and 100% pP support), but more about their relationships among each other can not be said according to this study. The Limoniidae genera *Scleroprocta* is grouped with the rest of the Limoniidae genera, only in the tree obtained from BI (pP = 0.87). The subfamily Limoniinae forms a monophyletic group,

according to combined data trees with 100% pP (BI) support, and moderate (MP), good (NJ) and strong (ML) bootstrap supports. Limoniinae is divided into two groups in the Bayesian tree, one of them containing genera *Limonia* and *Rhipidia* (pP = 1.00), and the other containing genera *Dicranomyia* and *Metalimnobia* (pP = 0.97). The ML tree grouped also *Limonia* and *Rhipidia* together with weak bootstrap support, but left the other genera unresolved. There can not be made any conclusion among the genera of Limoniinae according to MP and ML trees. The subfamily Limnophilinae is paraphyletic, according to the trees obtained in this study. Only the genera *Eloeophila* and *Phylidorea* are grouped together in Bayesian combined data tree (pP = 0.77). All trees grouped *Eloeophila maculata* and *E. trimaculata* species together (with strong bootstrap and pP values), but rest of the Limnophilinae species, including *Austrolimnophila unica*, *Paradelphomyia fuscula* and *Pilaria meridiana*, stayed unresolved.

4. DISCUSSION

This study was the first one that used molecular methods (mitochondrial and nuclear DNA) to infer phylogenetic relationships among Tipuloidea. The phylogenetic analyses of COI and Ef1α sequences confirmed the monophyly of the crane fly subfamilies, Tipulinae, Cylindrotominae, Uliinae, Pediciinae, Limoniinae and Chioneinae. The family Limoniidae and its subfamily Limnophilinae are paraphyletic according to this study.

The findings of the present study support in most part the results of the phylogenetic study by Oosterbroek & Theowald (1991), based on the morphological characters of larvae and pupae (see Fig. 1a). Both studies showed the paraphyly of the subfamily Limnophilinae and the monophyly of all the rest of the crane fly subfamilies. The Bayesian tree in this study clustered the genera of the subfamilies Pediciinae in the same way as the tree by Oosterbroek & Theowald (1991), In the Chioneinae *Molophilu*s was sister group of *Ormosia* and *Erioptera* basal to this group while in the tree by Oosterbroek & Teowald (1991) *Erioptera* was the sister group of *Molophilus* and *Ormosia* was basal to this group. On the contrary, the adult morphological character based hypothesis by Stary (1992) – in which Tipuloidea is divided into two groups; one containing family Limoniidae and the other including families Pediciidae, Tipulidae and Cylindrotomidae (see Fig. 1b) – did not get any supports from this study, as the Limoniidae subfamilies did not cluster together in this study.

The phylogenetic trees in this study gave new information concerning the Limoniinae genera *Rhipidia* and *Limonia*, which clearly form their own clade. In the morphology based tree by Oosterbroek & Theowald (1991), all the Limoniinae genera were left unresolved. Also the genera of the Cylindrotomidae were unresolved in the morphological tree, whereas in the present study, Cylindrotomidae genera *Diogma* and *Triogma* form a monophyletic group according to the ML, NJ and Bayesian trees.

The molecular trees clustered the Tipulidae subgenera *Tipula Savtshenkia* and *T. Vestiplex* as their own clades, supporting their taxonomic classification. Instead the separation of the genus *Dicranomyia* in the two subgenera *Dicranomyia* and *Numantis* was not supported by this study. However, more data are needed to confirm this later conclusion.

Most of the basal nodes of the phylogenetic trees obtained with different reconstruction methods show weak or moderate support according to the posterior probabilities or bootstrap, leaving unresolved most of the relationships among the crane fly damilies. Therefore the original purpose to infer the phylogenetic relationships among the four crane fly families was only partially obtained with the present material. The presence of saturation in the sequences and in particularly at the third position of the COI gene is

surely responsible for the lack of node support in the phylogenetic trees. COI is a faster evolving gene than nuclear Ef1α gene and its overall higher saturation level is not surprising. For this reason the nuclear Ef1α sequences gave more resolved trees than mitochondrial COI sequence data even at the lower branches of the tree, although it could not give reliable information on the relationships among the families or subfamilies either. COI gene is a recognized marker for species recognition in the barcoding of life (Hebert *et al.* 2003) and surely still useful for close analysis of relationships among species within genera.

Other reasons for this study failing to reveal more clearly the phylogenetic relationships among crane flies can be result of the short sequences that were used. Many phylogenetic studies have used longer fragments of the COI-COII and Ef1α genes, for example the studies concerning butterflies (Lepidoptera) with fragments of 1614-2169 bp COI-COII and 890-1195 bp Ef1α genes (Monteiro & Pierce 2001, Kandul *et al.* 2004, Silva-Brandão *et al.* 2005). Pereira *et al.* (2007) used totalling over 9000 bp of 8 different nuclear and mitochondrial DNA genes to construct phylogeny of pigeons and doves (Aves, Columbiformes). Initially I intended to sequence the whole Ef1α gene (1241 bp) and a longer fragment of the COI and COII gene complex (about 1899 bp). However, the primers used were not specially designed for crane flies and some did not work as planned. Thus I got only a small fragment of the COI gene (654 bp) and a fragment of 950 bp for the Ef1α gene. However, some studies have used smaller fragments of DNA, and still got reliable results. For example Gómez *et al.* (2002) reconstructed a phylogenetic tree of cryptic species complex *Brachionus plicatilis* (Rotifera) based on 957 bp of mitochondrial and nuclear gene fragments, and Mendel *et al.* (2008) used total of 1097 bp mitochondrial and nuclear sequences when studying the phylogenetic relationships among the ray-finned fish genus *Gobio* (Teleostei, Cyprinidae). Though, the species groups in these studies were closely related.

The present study clarifies the major internal relationships of crane flies and provides a useful data for future studies to reveal a more accurate phylogenety of this superfamily. Seen the differences between the morphological and molecular based trees, I would suggest that the family Limoniidae, and especially its subfamily Limnophilinae, should be taken for a closer inspection in future researches. Furthermore the relations between other tipuloidean subfamilies should be investigated in more detail with the new studies, including more species from all subfamilies, genera and their subgenera, specially the ones missing in this study. At least two individuals from each species should be added in the future studies. The sequences acquired from this study could be used to design specific primers for the Tipuloidea, which should give longer and better sequences for the analyses and more reliable results. New information sources could be used too, including different genes. For example study by Wheat & Watt (2008) used ribosomal DNA (rDNA) genes with mitochondrial COI and COII genes to construct phylogenetic tree of *Colius* butterflies (Lepidoptera, Pieridae) as the rDNA is considered to be useful when studying the deeper level of divergence (Simon *et al.* 1994). Also combined molecular and morphological data has been used in many phylogenetic studies successfully, for example with weevils (Coleoptera, Curculionoidea) (Marvaldi *et al.* 2002) and brush-footed butterflies (Lepidoptera, Nymphalidae) (Wahlberg & Nylin 2003). This was the first molecular based phylogeny that has really concentrated on relationships among the tipuloidean families, still there is much research to do.

ACKNOWLEDGEMENT

I would like to thank my instructors Prof. Alessandro Grapputo and Prof. Johanna Mappes, and the crane fly expert M.Sc Jukka Salmela for their help and advices. The staff of the DNA laboratory gave me a lot of help with my laboratory work, and deserves also my gratitude. The final thanks go to my family and friends, who supported me in many ways during my studies, even thought it sometimes felt like never-ending.

LITERATURE CITED

Alexander C. & Byers G. 1981. Tipulidae. In: McAlpine J., Peterson J., Shewell G., Teskey H., Vockeroth J. & Wood D. (eds), *Manual of Nearctic Diptera, Vol. 1*. Research Branch, Agriculture Canada, Monograph 27: 153–190.

Byers G. 1992. Crane flies – three families or one? *Acta zool. cracov.* 53: 37-41.

Cho S., Mitchell A., Regier J., Mitter C., Poole R., Friedlander T. & Zhao S. 1995. A highly conserved nuclear gene for low-level phylogenetics: *Elongation factor-1α* recovers morphology-based tree for Heliothine moths. *Mol. Biol. Evol.* 12: 650-656.

Clary D. & Wolstenholme D. 1985. The mitochondrial-DNA molecule of *Drosophila yakuba* – nucleotide-sequence, gene organization, and genetic-code. *J. Mol. Evol.* 22: 252-271.

Culver M., Johnson W., Pecon-Slattery J. & O'Brien S. 2000. Genomic ancestry of the American puma (*Puma concolor*). *J.Hered.* 91: 186-197.

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

Daugherty C., Cree A., Hay J. & Thompson M. 1990. Neglected taxonomy and continuing extinctions of tuatara (*Sphenodom*). *Nature* 347: 177-179.

de Jong H. 1994. The phylogeny of the subgenus *Tipula* (*Savtshenkia*) (Diptera. Tipulidae), with special reference to the western Mediterranean fauna. *Tijdschrift voor Entomologie* 137: 271-323.

de Jong H. 1995. The phylogeny of the *Tipula* (*Lunatipula*) *bullata* and *falcata* species groups (Diptera: Tipulidae). *Tijdschrift voor Entomologie* 138: 245-267.

de Jong H., Oosterbroek P., Gelhaus J., Reusch H. & Young C. 2008. Global diversity of craneflies (Insecta, Diptera: Tipulidea or Tipulidae *sensu lato*) in freshwater. *Hydrobiologia* 595:457–467.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.

Foley D., Wilkerson R., Cooper R., Volovsek M. & Bryan J. 2007. A molecular phylogeny of *Anopheles annulipes* (Diptera: Culicidae) sensu lato: The most species-rich anopheline complex. *Molecular phylogenetics and evolution* 43: 283-297.

Folmer O., Black M., Hoeh W., Lutz R. & Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molec. Mar. Biol. Biotechnol.* 3: 294-299.

Fox G. & 18 others. 1980. The phylogeny of prokaryotes. *Science* 209: 457-463.

Frankham R., Ballou J. & Briscoe D. 2002. *Introduction to conservation genetics.* Cambridge University Press, UK, 617 p.

Friedrich M. & Tautz D. 1997. Evolution and phylogeneny of the Diptera: a molecular phylogenetic analysis using 28s rDNA sequences. *Syst. Biol*. 46: 674-698.

Gómez A., Serra M., Carvalho G. & Lunt D. 2002. Speciation in ancient cryptic species complexes: Evidence from the molecular phylogeny of *Brachionus plicatilis* (Rotifera). *Evolution* 56: 1431–1444.

Graur D. & Li W. 2000. *Fundamentals of molecular evolution.* Sinaeur Associates, Inc., U.S.A., 481 p.

Guindon S., Lethiec F., Duroux P. & Gascuel O. 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33: 557-559.

Hebert P., Penton E., Burns J., Janzen D. & Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *PNAS* 101: 14812–14817.

Hebert P., Ratnasingham S. & deWaard J. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B* 270: 96–99.

Huelsenbeck J. and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17: 754-755.

Kandul N., Lukhtanov V., Dantchenko A., Coleman J., Sekercioglu C., Haig D. & Pierce N. 2004. Phylogeny of *Agrodiaetus* Hübner 1822 (Lepidoptera: Lycaenidae) Inferred from mtDNA sequences of *COI* and *COII* and nuclear sequences of *EF1-α*: Karyotype diversification and species radiation. *Syst. Biol.* 53: 278-298.

Lin C. & Danforth B. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ. Insights from Bayesian analyses of combined datasets. *Mol. Phylogenet. Evol.* 30: 686-702.

Marvaldi A., Sequeira A., O'Brien C. & Farrell B. 2002. Molecular and morphological phylogenetics of weevils (Coleoptera, Curculionoidea): Do niche shifts accompany diversification? *Syst. Biol.* 51: 761-785.

Mendel J., Lusk S., Vasil'eva E., Vasil'ev V., Lusková V., Ekmekci F., Erk'akan F., Ruchin A., Koščo J., Vetešník L., Halačka K., Šanda R., Pashkov A. & Reshetnikov S. 2008. Molecular phylogeny of the genus *Gobio* Cuvier, 1816 (Teleostei: Cyprinidae) and its contribution to taxonomy. *Molecular Phylogenetics and Evolution* 47: 1061–1075.

Monteiro A. & Pierce N. 2001. Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from COI, COII, and EF-1a gene sequences. *Molecular phylogenetics and evolution* 18: 264–281.

Nuin P. 2008. MrMTgui: cross-platform interface for ModelTest and MrModeltest. http://www.genedrift.org/mtgui.php.

Nylander, J. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Oosterbroek P. 2008. Catalogue of the Craneflies of the World. http://nlbif.eti.uva.nl/ccw/index.php.

Oosterbroek P. & Courtney G. 1995. Phylogeny of the Nematocerous families of Diptera (Insecta). *Zoological Journal of the Linnean Society* 115: 267-311.

Oosterbroek P. & Theowald B. 1991. Phylogeny of the Tipuloidea based on characters of larvae and pupae (Diptera, Nematocera). *Tijdschrift voor Entomologie* 134: 211-267.

Pereira S., Johnson K., Clayton D. & Bakeri A. 2007. A mitochondrial-DNA-based phylogeny for some evolutionary-genetic model species of *Colias* butterflies (Lepidoptera, Pieridae). *Syst. Biol.* 56: 656–672.

Posada D. & Crandall K. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Roca A., Georgiadis N., Pecon-Slattery J. & O'Brien S. 2001. Genetic evidence for two species of elephant in Africa. *Science* 293: 1473 – 1477.

Rodrígues F., Oliver J., Marín A. & Merina J. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142: 485-501.

Ruvolo M. 1997. Molecular phylogeny of the hominoids: Inferred from multiple independent DNA sequence data sets. *Mol. Biol. Evol.* 14: 248-265.

Sarich V. & Wilson A. 1967. Immunological time scale for hominid evolution. *Science* 158: 1200-1203.

Silva-Brandão K., Freitas A., Brower A. & Solferini V. 2005 Phylogenetic relationships of the New World Troidini swallowtails (Lepidoptera: Papilionidae) based on COI, COII and EF-1α genes. *Molecular Phylogenetics and Evolution* 36: 468-483.

Simon C., Frati F., Beckenback a., Cresp B., Liv H. & Flook P. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. *Ann. Entomol. Soc. Am.* 87: 651-701.

Stary J. 1992. Phylogeny and classification of Tipulomorpha, with special emphasis on the family Limoniidae. *Acta zool. cracov.* 35: 11-36.

Swofford D. 2003. PAUP*. Phylogenetic analysis using parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Wahlberg N. & Nylin S. 2003. Morphology versus molecules: resolution of the positions of Nymphalis, Polygonia, and related genera (Lepidoptera: Nymphalidae). *Cladistics* 19: 213-223.

Wheat C. & Watt W. 2008. A mitochondrial-DNA-based phylogeny for some evolutionary-genetic model species of *Colias* butterflies (Lepidoptera, Pieridae). *Molecular Phylogenetics and Evolution* 47: 893–902.

Woese C. & Fox G. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Acad. Sci. USA* 74: 5088-5090.

Woese C., Kandler O. & Wheelis M. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87: 4576-4579.

Wood D. & Borkent A. 1989. Phylogeny and classification of the Nematocera. In: McAlpine J. & Wood M. (eds.) *Manual of Nearctic Diptera, Vol. 3.* Research Branch, Agriculture Canada, Monograph 32: 1333-1370.

Xia X. & Xie Z. 2001. DAMBE: Data analysis in molecular biology and evolution. *Journal of Heredity* 92: 371-373.

Xia X., Xie Z., Salemi M., Chen L. & Wang J. 2003. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* 26: 1-7.

Yeates D., Wiegmann B., Courtney G., Meier R., Lambkin C. & Pape T. 2007. Phylogeny and systematics of Diptera: two decades of progress and prospects. *Zootaxa* 1668:565-590.