Molecular systematics of the avian superfamily Sylvioidea with special regard to the families Acrocephalidae and Locustellidae (Aves: Passeriformes)

I n a u g u r a l d i s s e r t a t i o n

zur

Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

an der Mathematisch-Naturwissenschaftlichen Fakultät

der

Ernst-Moritz-Arndt-Universität Greifswald

vorgelegt von

Silke Fregin

geboren am 22.05.1976

in Werneck

Greifswald, Juni 2012

Dekan:	Prof. Dr. Klaus Fesser
1. Gutachter :	Prof. Dr. Gabriele Uhl
2. Gutachter:	Prof. Dr. Staffan Bensch
Tag der Promotion:	14.11.2012

Table of contents

Zusammenfassung

Das Ziel der Arbeit war es, die systematischen Verwandtschaftsbeziehungen innerhalb der Überfamilie der Sylvioidea (Aves: Passeriformes), im Allgemeinen und innerhalb der nah verwandten Familien Acrocephalidae und Locustellidae im Speziellen, anhand von DNA-Sequenzen zu untersuchen.

Sylvioidea (Grasmückenartige) selbst, und die zugehörigen Familien waren bereits Fokus von zahlreichen Untersuchungen, basierend sowohl auf morphologischen Merkmalen, als auch auf der Ebene von DNS. Aufgrund ihrer morphologischen Ähnlichkeit und der vermutlich zeitlich schnellen Ausbreitung, haben es die meisten Studien nicht geschafft die Verwandtschaftsverhältnisse zwischen den Familien innerhalb der Grasmückenartigen aufzulösen. Auch die systematische Abgrenzung der einzelnen Familien und die Beziehungen der zugehörigen Arten untereinander sind immer noch nicht komplett gelöst. In der vorliegenden Untersuchung wurden sowohl größere Datensets in Bezug auf Anzahl der Taxa und DNS-Sequenzen, als auch differenziertere Methoden zur Analyse herangezogen, um eine bessere Auflösung der Verwandtschaftsbeziehungen innerhalb der Sylvioidea, Acrocephalidae und Locustellidae zu erzielen. Darüber hinaus wurde die Anwendbarkeit von Barcoding für die Familie Acrocephalidae getestet.

Die Monophylie der Sylvioidea konnte bestätigt werden und die Familien Paridae und Remizidae, welche manchmal zu den Grasmückenartigen gezählt werden, befanden sich zwischen den Taxa der Außengruppe. Vier Familien, Nicatoridae, Panuridae, Alaudidae und Macrosphenidae bilden die ersten Splits innerhalb der Sylvioidea. Die Aufteilung der früheren Gruppe aus Sylviiden und Timaliiden in fünf Familien Sylviidae, Leiothrichidae, Pellorneidae, Timaliidae, und Zosteropidae konnte bestätigt werden. Es wurde gezeigt, dass *Scotocerca*, *Erythrocercus* und *Hylia* die kürzlich vorgeschlagen wurden Mitglieder der Cettiidae zu sein, nicht zu dieser Familie gehören. Aufgrund ihrer morphologischen und ökologischen Verschiedenheit wurde empfohlen diese drei Gattungen jeweils zur Familie zu erheben, Scotocercidae, Erythrocercidae und Hyliidae.

Die Familie der Acrocephalidae bestand aus vier Gattungen: *Nesillas*, *Acrocephalus*, *Hippolais*, und *Chloropeta*. In den Analysen zeigte sich, dass die letzten drei Gattungen nicht monophyletisch sind. Eine Art der Gattung *Acrocephalus*, *A. aedon*, war Schwester zu einer Gruppe bestehend aus vier *Hippolais*-Arten und zwei von drei *Chloropeta*-Arten. Diese Gruppen wurden unter dem Gattungsnamen *Iduna* zusammengefaßt, aufgrund der DNS Analysen und gemeinsamer morphologischer und ökologischer Merkmale. Dem „International Code of Zoological Nomenclature“ entsprechend, hat der Name *Iduna* Priorität gegenüber *Hippolais* oder *Chloropeta*. Die eine verbleibende *Chloropeta* Art (*C. gracilirostris*) musste in *Calamonastides* umbenannt werden, da *Chloropeta* aufgrund

der Nomenklaturregeln nun nicht mehr zur Verfügung stand.

In die Analyse der Familie Locustellidae wurden sieben Gattungen einbezogen: *Locustella*, *Bradypterus, Megalurus*, *Dromaeocercus*, *Schoenicola*, C*incloramphus* und *Eremiornis*. Abgesehen von den monotypischen Gattungen *Dromaeocercus* und *Eremiornis* und *Schoenicola,* von der nur eine Art enthalten war, waren die übrigen Gattungen nicht monophyletisch. Eine Gruppe beinhaltete alle *Locustella*-Arten, *Megalurus pryeri* und alle asiatischen bzw. orientalischen *Bradypterus*-Arten. Diese Gruppe wurde komplett in *Locustella* umbenannt, da hier die Typus-Art von *Locustella* enthalten ist, die von *Bradypterus* hingegen in eine andere Gruppe fiel. Deshalb behielten die übrigen afrikanischen *Bradypterus*-Arten ihren Gattungsnamen, und *Dromaeocercus*, ebenfalls phylogenetisch zu dieser Gruppe gehörig, wurde in *Bradypterus* umbenannt. *Cincloramphus*, der eine gemischte Gruppe mit den weiteren *Megalurus* Arten bildet, wurde mit dieser synonomisiert.

Barcoding, eine zunehmend beliebte Methode Arten abzugrenzen, wurde auf ihre Anwendbarkeit für die Familie Acrocephalidae getestet. Vierzehn Taxa, die gegenwärtig vollen Artstatus besitzen, würden unter den empfohlenen Schwellenwert von 2% Sequenzdivergenz fallen, der für die Abgrenzung von Arten gilt, wenn man das mitochondriale Cytochrom b Gen heranzieht. Es wurde auch gezeigt, dass es wichtig ist darzulegen, welcher Abschnitt einer DNS-Sequenz verwendet wird, da verschiedene Bereiche einer Sequenz unterschiedliche Ergebnisse bezüglich der 2% Schwelle liefern können. Außerdem ist es bei unvollständigen Sequenzen wichtig die Wahl zwischen „Kompletter Deletion“ oder „Paarweiser Deletion“ bei der Berechnung von genetischen Distanzen zu haben.

Summary

The goal of this thesis was to study the systematic relationships within the superfamily Sylvioidea (Aves: Passeriformes) in general and within the closely related families Acrocephalidae and Locustellidae in particular, by means of DNA sequences.

Sylvioidea itself and families therein were the focus of many studies based as well on morphological characters as on DNA. Due to their morphological similarity and their presumably rapid radiation most studies failed to solve relationships between sylvioidean families and also demarcations of single families and relations within are still in progress. In this study, an enlargement of previous datasets, both taxa and number of DNA sequences, and more sophisticated analysis methods were used to improve the resolution in Sylvioidea, Acrocephalidae and Locustellidae. In addition, the applicability of barcoding in Acrocephalidae was tested.

The monophyly of Sylvioidea could be supported and the families Paridae and Remizidae, which were sometimes still included, clustered among the outgroup taxa. Four families, Nicatoridae, Panuridae, Alaudidae, and Macrosphenidae constitute basal splits within Sylvioidea. The division of the former sylviid/timaliid clade in five families, Sylviidae, Leiothrichidae, Pellorneidae, Timaliidae, and Zosteropidae was supported. *Scotocerca*, *Erythrocercus*, and *Hylia*, previously supposed to be members of Cettiidae, were shown not to belong to this family. As the three genera are also morphologically and ecologically different, they were here proposed to be elevated to family rank, Scotocercidae, Erythrocercidae and Hyliidae, respectively.

The family Acrocephalidae consisted of the four genera, *Nesillas*, *Acrocephalus*, *Hippolais*, and *Chloropeta*. In the analysis for this thesis, the latter three appeared to be non-monophyletic. One *Acrocephalus* species, *A. aedon* was sister to a clade containing four species of *Hippolais* as well as two out of three *Chloropeta* species. They were all merged in the genus *Iduna*, based on the DNA evidence and shared morphological and ecological characters. *Iduna* had priority over *Hippolais* or *Chloropeta* according to the International Code of Zoological Nomenclature. The one remaining *Chloropeta* species (*C. gracilirostris*) had to be renamed to *Calamonastides* as *Chloropeta* was no longer available for this taxon.

Seven genera were included in the re-analysis of the family Locustellidae: *Locustella*, *Bradypterus, Megalurus*, *Dromaeocercus*, *Schoenicola*, C*incloramphus*, and *Eremiornis.* Apart from the monotypic genera *Dromaeocercus* and *Eremiornis* and *Schoenicola*, of which only one species was included, the remaining genera were found to be non-monophyletic. One clade contained all *Locustella* species,

Megalurus pryeri and all Asian/Oriental *Bradypterus* species. All species in this clade were synonymized with *Locustella*, as the type species of *Locustella* was included, whereas the type species of *Bradypterus* fell in a different clade. Therefore, the remaining African *Bradypterus* species retained their genus name, and *Dromaeocercus* was renamed to *Bradypterus* as it clustered within *Bradypterus*. *Cincloramphus*, intermingling with the remaining *Megalurus* species, was synonymized with the latter.

Barcoding, growing in popularity for delimiting species, was tested in its applicability for Acrocephalidae. Fourteen taxa currently recognized as full species would fall under the 2% threshold of sequence divergence proposed for delimiting species using the mitochondrial cytochrome b gene. It was also shown that it is important to clarify which part of a DNA sequence is used, as different parts can give different results regarding the 2% threshold. In addition, the choice of "complete deletion" or "pairwise deletion" in calculating genetic distances is important, if incomplete are sequences used.

Abbreviations

BI	Bayesian inference
bp	basepairs
C	cytosine
cm	centimetre
CO1	cytochrome oxidase 1
CO2	cytochrome oxidase 2
DNA	deoxyribonucleic acid
FGB	fibrinogen beta chain
g	gram
G	guanine
GAPDH	glyceraldehyd–3–phosphat dehydrogenase
GTR	general time reversible
indel	insertion and/or deletion
IOC	International Ornithological Congress
JC	Jukes and Cantor
LDHB	lactate-dehydrogenase B
MB	myoglobin intron
ML	maximum likelihood
MP	maximum parsimony
MT–CYB	mitochondrially encoded cytochrome b
mtDNA	mitochondrial deoxyribonucleic acid
N	unknown nucleotide
ND2	NADH-ubiquinone oxidoreductase chain 2
numt	nuclear mitochondrial DNA
ODC1	ornithine-decarboxylase 1
PCR	polymerase chain reaction
PP	posterior probability
RAG1	recombination activation gene 1
RAG2	recombination activation gene 2
ssp	subspecies
12S	mitochondrially encoded 12S RNA

1 General introduction

1.1 DNA-DNA hybridisation studies

The modern (r)evolution of bird systematics started with the work of Charles G. Sibley and Jon E. Ahlquist in the 1980s which was summarized in: "Phylogeny and Classification of Birds – A Study in Molecular Evolution" (1990), also called "The Tapestry" (Gill and Sheldon 1991, Barker et al. 2002, Dyke and Kaiser 2011). Their DNA-DNA hybridisation studies were based on the concept that DNA-strands will separate if heated, because hydrogen bonds between pairing nucleotides are dissociated (melting). Dissociated strands can then be mixed with likewise treated DNA from different species. Cooling down the temperature to ~60°C allows rebonding of complementary homologous strands, also between different species, thus producing DNA-hybrids. These hybrid double-strands have different thermal stabilities when heated again, corresponding to their degree of relatedness (or rather similarity). The more mutations have accumulated over time the more mismatches between nucleotides of complementary DNA-strands will occur and thus less energy is necessary to melt these strands again (Sibley and Ahlquist 1982). From these characteristic melting-curves genetic distances can be calculated and a phylogeny inferred. In this way Sibley and Ahlquist built a new systematic framework for all birds, which was subsequently implemented in "Distribution and Taxonomy of birds of the world" by Sibley and Monroe (1990). It was the first taxonomy for all birds not based on morphological data, which are much more prone to erroneous conclusions due to convergent evolution (e.g. Mayr 2011). Convergently similar or even identical characteristics have its cause not in close relatedness, but in adaptation to a similar environment (Freeman and Heron 2001).

As DNA sequencing became standard, several of the taxa established by Sibley and Ahlquist (1990) and Sibley and Monroe (1990) were found to be non-monophyletic (e.g. Barker et al. 2002, Ericson and Johansson 2003, Alström et al. 2006a). A monophylum is a group or clade that contains all descendants of a common ancestor. If a set of species includes a common ancestor and some, but not all, of its descendants, it is called paraphyletic (Freeman and Herron 2001). A polyphyletic group contains a set of species, which are not descendant from a common ancestor, but grouped by similarity (Freeman and Herron 2001). Thus, para- and polyphyletic taxa are considered non-natural and should therefore not be named.

These erroneous classifications, to our present knowledge, could be due to problems which are connected with DNA-DNA hybridisation. Mindell (1992) summarized some difficulties that can occur. First, it is problematic to separate the single-copy DNA from repetitive parts and multi-copy genes. This is necessary to avoid paralogous hybrid DNA, i.e. binding of a duplicated gene with its original. Only orthologous or homologous genes, which are related by speciation and not by duplication, show

the true genetic divergence between two species. A single-copy genome has also presumably the advantage of the same size and an equal spatial arrangement of the sequences, thus assuring binding of only orthologous sequences over the whole DNA fragment. Repeated hybridisation experiments may show a higher variation than that found between species, making analysis between closely related species difficult. Also relationships of very distantly related taxa are difficult to analyse as their melting curves cannot be observed directly due to the lack of homologous sequences and must therefore be extrapolated.

1.2 Advantages and disadvantages of DNA sequences

DNA sequences are in many respects advantageous compared to DNA-DNA hybridisation. It is possible to use specific parts of the genome, mitochondrial and/or nuclear, chosen for a particular purpose of a study. Fast evolving parts, like mitochondrial sequences (mtDNA) or nuclear introns, are suitable for resolving young evolutionary relationships, e.g. between species. Whereas older relationships are better analysed with more conservative genes like nuclear exons (e.g. Lin and Danforth 2004). Mostly, a combination of both is used to target different parts of a phylogenetic tree, respectively. With the improvement of phylogenetic inference programs it is now possible to implement different substitution models for different parts of an alignment, e.g. different genes or codon positions.,thus, taking into account that distinct parts of a genome or gene evolve differently (Wolfe et al. 1989, Yang 1995, Roe and Sperling 2007). The advantage of the DNA sequence is the possibility to compare the actual mutations that have occurred, instead of the indirect way of comparing melting curves. The simplest way to do so is counting the number of differences between any pair of sequences. The more such substitutions have occurred, the more distantly related are the organisms or taxa compared.

But molecular systematics has its drawbacks as well, e.g. numts (nuclear mitochondrial DNA: insertions of mtDNA into the nuclear genome), long-branch attraction, or saturation effects may obscure true phylogenetic relationships.

Nuclear copies of mitochondrial DNA pieces were known since the 1980s from different taxa (fungi: Wright and Cummings 1983, maize: Kemble et al. 1983, locusts: Gellissen et al. 1983, snow geese: Quinn and White 1987). The term numt was introduced later by Lopez et al. (1994). Once transferred to the nucleus the copy of the mitochondrial sequence evolves independently from its original sequence. Differences between the mitochondrial original and the nuclear copy from < 1% in aphids (Sunnocks and Hales 1996) to >25% in primates (Collura and Stewart 1995) have been reported. However, the nuclear copy seems to evolve more slowly than their mitochondrial origin, thus resembling more the ancestral state of the mitochondrial nucleotide sequence (Zhang and Hewitt

1996). These molecular "fossils" pose not only a risk, but also have the potential for alternative approaches for evolutionary studies. To avoid sequencing numts it is possible to separate mtDNA from the nuclear DNA content of the blood or tissue sample. Primers designed for a large number of very different taxa on a consensus basis can be problematic, as these potentially resemble the ancestral nucleotide sequence and thus could favour sequencing numts instead of the mtDNA. Also, larger mitochondrial PCR (polymerase chain reaction) products are preferable over smaller fragments, because the risk of amplifying numts is reduced in larger products assuming that shorter fragments are more likely to be copied into the nucleus than larger ones (Sorenson and Quinn 1998).

Long-branch attraction can occur when two or more taxa rapidly accumulate mutations resulting in long branches in a phylogenetic tree. As there are only a limited number of nucleotide states (four, for the four nucleotides), there is a certain probability that the sequences become similar over time and thus will cluster in an analysis, although they are not close relatives (Felsenstein 1978, Hendy and Penny, 1989). But this problem can sometimes be overcome adding more closely related taxa, "breaking" the long branch into shorter ones (Anderson and Swofford 2004, Bergsten 2005). Saturation effects have the same basis. As more and more substitutions take place at any particular site across the entire sequence, the phylogenetic signal gets lost as the sequences become similar because of chance alone (Hackett 1996, Griffiths 1997). This can already be problematic for the alignment of sequences. In an alignment all sequences are aligned so that homologous characters (positions) are written in one column. If sequences are very diverse, homologous nucleotides may be difficult to identify. Also the arrangements of insertions or deletions (indels), thus producing gaps in the alignment, can cause difficulties. As different alignments based on the same sequences can influence the topology of a phylogenetic tree (Morrison and Ellis 1997), great care should be taken in the process of constructing an alignment.

DNA sequences have their assets and drawbacks, just like any other character type such as DNA-DNA hybridisation (Sarich et al. 1989, Springer and Krajewski 1989) and morphological characters (e.g. Livezey and Zusi 2007, Mayr 2011). Therefore, it would be preferable to use more inclusive approaches including more than one set of characters (Griffiths et al. 2004, Bartelli and Giannini 2005, Haase et al. 2007, Alström et al. 2008, Capa et al. 2010, Ekrem et al. 2010) to critically assess results from different perspectives.

1.3 Analysis of DNA sequences

Basically, there are four commonly used methods to infer a phylogenetic tree from DNA sequences and the substitutions which have taken place: 1) Distance methods, 2) Maximum Parsimony, 3) Maximum Likelihood, 4) Bayesian Inference.

Distance methods use DNA sequences only indirectly in calculating first the genetic distances between every pair of sequence and then inferring a phylogenetic tree based on the resulting distance matrix. In contrast, all other methods use the sequences directly and are therefore called character-based methods. Maximum Parsimony (MP) searches for the phylogenetic tree were the smallest number of evolutionary changes are required to explain the observed data (Kluge and Farris 1969). Whereas Maximum Likelihood (ML) calculates the likelihood for every possible tree and tries to find the tree, that explains the observed data with the highest probability, thus the tree with the highest likelihood. But the likelihood of a tree does not equate with the probability that it is the correct tree (Felsenstein 1981). Bayesian inference (BI) uses Markov chain Monte Carlo simulation techniques to find the tree with the maximum posterior probability (PP). PPs can be interpreted as the probability that the tree is correct (Huelsenbeck et al. 2001), contrary to the likelihood values of the maximum likelihood analysis. Posterior probabilities are obtained by combining the prior probability of a phylogeny (Pr[Tree]) with the likelihood of the sequence alignment conditional on a tree (Pr[Data|Tree]) (Bayes´s theorem, Huelsenbeck et al. 2001). The prior probability is typically assumed to be equal between all trees, the likelihood and thus the probability to observe the sequence alignment requires making specific assumptions, like choosing an appropriate evolutionary substitution model. Substitution models are necessary to account for the stochastic process of multiple mutations at certain nucleotide sites (Swofford et al. 1996). In total, there are four parameters which can be considered in nucleotide substitution models: 1) the substitution rate among nucleotides, 2) the frequency of the four nucleotides, 3) the amount of invariable sites, and 4) the heterogeneity of substitution rates among sites (Posada and Crandall 2001). These parameters can be differently combined or allowed to vary, thus resulting in different specific models. The simplest model is known as the Jukes-Cantor model (JC; Jukes and Cantor 1969), where it is assumed that all nucleotides occur at the same frequency and all nucleotides have the same probability to change into any of the three others, only the mean rate of substitution can vary for different datasets. The most complex of the commonly implemented models is the General Time Reversible (GTR) model (Lanave et al. 1984, Tavaré 1986, Rodríguez et al. 1990), which has a symmetrical substitution matrix. This means that e.g. G (Guanine) is substituted by a C (Cytosine) with the same rate as C by a G. Between JC and GTR jModeltest (Posada 2008), a program used for model-fitting, currently offers another 86 models.

1.4 Possibilities beyond phylogeny

DNA sequences are not only used for solving relationships in a phylogenetic tree, like between e.g. species, genera, families etc., but also have a wide range of other applications in population genetics (Powell et al. 1995; Fernando et al. 2000, Steane et al. 2006, Backström et al. 2008), dating of

evolutionary events (Hasegawa et al. 1985, Garnery et al. 1992, Aguileta et al. 2006, Jiang et al. 2009), delimiting species and detecting cryptic species (Collins and Paskewitz 1996, Helbig et al. 1996, Bradley and Baker 2001) or identifying species from (unknown) DNA samples (Hugenholtz and Pace 1996, Mackie et al. 1999, Hebert et al. 2003, Dalén et al. 2004). For the identification of organisms, DNA sequences were first used for microorganisms, to screen environmental samples and assess their diversity (e.g. Nanney 1982, Pace et al. 1986). Two decades later Hebert et al. (2003a, 2003b) suggested to use a ~650bp (basepairs) fragment of the CO1 gene (cytochrome oxidase 1) for a uniform bioidentification system for all animals. This genetic "barcode" is based on the prerequisite that sequence divergence within species does not exceed 2% and divergence between species is ten times higher, leaving a "barcoding gap" inbetween. While delimiting species by means of DNA sequences was occasionally done before (e.g. Wesson et al. 1993, Klicka et al. 1999, Burbrink et al. 2000, Hickerson et al. 2003), after introducing barcoding the number of such studies increased dramatically (e.g. Greenstone et al. 2005, Saunders 2005, Ward et al. 2005, Clare et al. 2007). But barcoding does not only have proponents. The approach has been criticized for various reasons, e.g. because no barcoding gap was found (Meier et al. 2006, Davison et al. 2009, Langhoff et al. 2009), that it is only an artefact of poor taxon sampling (Wiemers and Fiedler 2007) or that using mean distances produces artificially a barcoding gap, what would not happen if always the smallest distances would be utilized (Meier et al. 2008). Also numts may be a risk, because of the small barcoding fragment (Song et al. 2008). Thus, there are many aspects to consider and examine before barcoding should be utilized in general for a certain study group.

2 Introduction to Sylvioidea

2.1 Systematic classification of the Sylvioidea

"...passerine phylogenies look like an upended head of an artist´s camel hair paintbrush with the myriad single strands inextricably mixed." (Feduccia, 1996)

Passeriformes (perching birds or Passerine birds) is the largest order among birds. They are divided into Oscines ("true" songbirds) and Suboscines (Tyranni) with Acanthisitti (New Zealand wrens) as their common sister-group (Fig. 1).

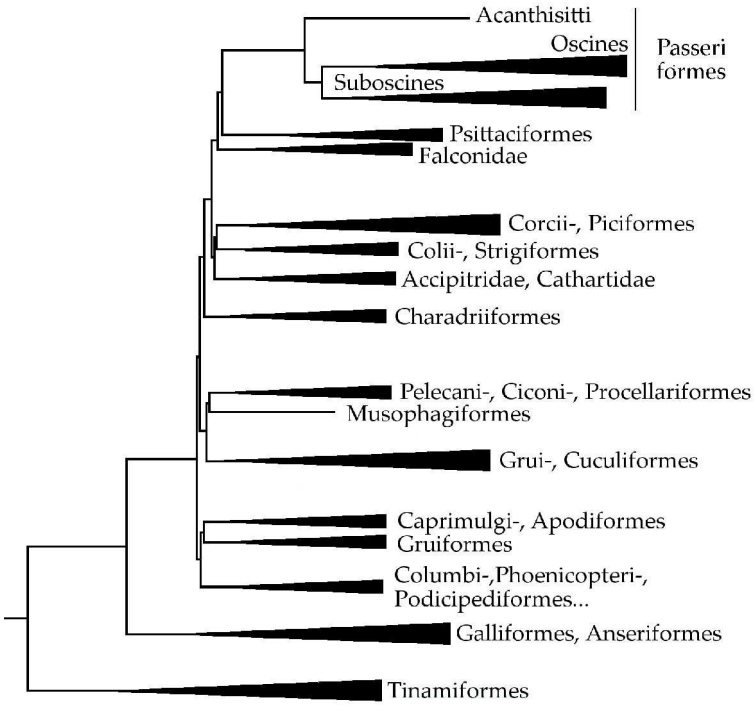

Figure 1: Tree of the class aves, derived and simplified after Hackett et al. (2008).

The suborder Acanthisitti, endemic to New Zealand, consists of only one family (Acanthisittidae) with two extant species in two genera. Three others went extinct, probably after the Maori started to colonize New Zealand (Gill 2004). Suboscines, or Tyranni, constitute a monophyletic group based on morphological synapomorphies, i.e. is the anatomy of the syrinx (Forbes 1882, Ames 1971) and the middle-ear ossicle (Feduccia 1974), as well as molecular data (Sibley and Ahlquist 1990, Irestedt et al. 2001, Chesser 2004). Oscines or Passeri are a unique group in so far as they are characterized by extreme morphological uniformity, thus making it very difficult to define subgroups based on structural characters that have often evolved by convergence rather than by common descent (Feduccia 1996). This complication has characterized the history of classification within oscines. The

number of recognized families varied from two (Fürbringer 1888) to 104 (Wolters 1975-82). Sibley and Ahlquist (1990) split the passerines in 45 families based on their DNA-DNA hybridisation studies, whereas Gill and Donsker (2011) recognized 106 families, taking recent studies into account. Sibley and Ahlquist (1990) divided the true songbirds in two parvorders, Corvida and Passerida. While the latter where confirmed as being monophyletic, Corvida where not (Ericson et al. 2000, Barker et al. 2002, 2004, Ericson et al. 2002). Passerida was divided into three superfamilies by Sibley and Ahlquist (1990): Muscicapoidea, Sylvioidea and Passeroidea, but considerable classificatory changes have taken place in the last two decades. Sylvioidea consisted of 11 families: Sittidae, Certhiidae, Paridae, Aegithalidae, Hirundinidae, Regulidae, Pycnonotidae, Hypocoliidae (incertae sedis), Cisticolidae, Zosteropidae, Sylviidae. Sittidae and Certhiidae were shown to fall outside Sylvioidea (Barker et al. 2002, Beresford 2005) and were sometimes considered as one superfamily (Certhioidea: Cracraft et al. 2004, Johansson et al. 2008). Regulidae and Hypocoliidae also were shown not to be part of Sylvioidea (e.g. Barker et al. 2002, Cibois and Cracraft 2004, Spellman et al. 2008). Paridae have a conflicting role, sometimes they were found outside Sylvioidea (e.g. Barker et al. 2002, Ericson and Johansson 2003, Alström et al. 2006a, Fuchs et al. 2006) and sometimes they were included (e.g. Sangster et al. 2010, Clements et al. 2011). Alaudidae, included in Passeroidea by Sibley and Ahlquist (1990) and Sibley and Monroe (1990), where later recovered as a sylvioid taxa (Sheldon and Gill 1996, Groth 1998, Barker et al. 2002). With the addition of more species and genetic markers to individual analyses species composition and phylogeny of families became clearer (e.g Cibois et al. 2001, Cibois 2003, Sheldon et al. 2005, Olsson et al. 2005, Moyle and Marks 2006, Nguembock et al. 2007, Gelang et al. 2009, Päckert et al. 2010, Alström et al. 2011a). Within Sylvioidea, not only higher level relationships changed but also the composition of families, partly considerably (Table 1). However, relationships between families and even superfamilies are still largely unresolved.

Table 1: Affiliation of genera to families within Sylvioidea in the latest printed classification of Dickinson (2003) compared to the online based IOC World bird names list version 2.10 (Gill and Donsker 2011). Genera included in this thesis are given in bold. Genera in parenthesis in the right column indicate the former genus name of renamed and/or splitted taxa. *: genera not included in Sylvioidea anymore.

Sylvioidea	Dickinson 2003	Sylvioidea	IOC World Bird Names List 2.10
		1 Panuridae	*Panurus*
		2 Nicatoridae	*Nicator*
		3 Alaudidae	***Mirafra, Ammomanes, Alauda*** *Heteromirafra, Calendulauda, Certhilauda, Pinarocorys, Chersomanes, Alaemon, Ramphocoris, Melanocorypha, Calandrella, Spizocorys, Eremalauda, Chersophilus, Pseudalaemon, Galerida, Lullula, Eremopterix, Eremophila*
		4 Pycnonotidae	***Pycnonotus, Arizelocichla*** **(***Andropadus***),** ***Atimastillas*** **(***Chlorocichla***),** ***Phyllastrephus, Hypsipetes*** **(***Ixos),*

Sylvioidea	Dickinson 2003	Sylvioidea	IOC World Bird Names List 2.10
			Spizixos, Stelgidillas, Eurillas, Calyptocichla, Baeopogon, Ixonotus, Chlorocichla, Thescelocichla, Bleda, Criniger, Alophoixus, Acritillas, Setornis, Tricholestes, Iole, Ixos, Thapsinillas, Hemixos, Cerasophila, Neolestes, Malia
Hirundinidae	Pseudochelidoninae: *Pseudochelidon* Hirundininae: ***Hirundo***[5], ***Delichon***[5] *Psalidoprocne, Pseudhirundo, Cheramoeca, Phedina, Riparia, Tachycineta, Progne, Pygochelidon, Notiochelidon, Haplochelidon, Atticora, Neochelidon, Stelgidopteryx, Alopochelidon, Ptyonoprocne, Cecropis, Petrochelidon*	**5** Hirundinidae	***Hirundo, Delichon,*** *Pseudochelidon, Psalidoprocne, Pseudhirundo, Cheramoeca, Phedina, Riparia, Tachycineta, Progne, Notiochelidon, Haplochelidon, Atticora, Neochelidon, Stelgidopteryx, Alopochelidon, Ptyonoprogne, Cecropis, Petrochelidon*
		6 Pnoepygidae	***Pnoepyga***
		7 Macrosphenidae	***Melocichla, Sphenoeacus, Macrosphenus, Sylvietta, Cryptillas*** *(Bradypterus)* *Achaetops*
		8 Cettiidae	***Scotocerca, Tesia, Cettia, Abroscopus, Erythrocercus, Hylia,*** *Hemitesia,Urosphena, Tickellia, Graueria, Phyllergates, Pholidornis*
Aegithalidae	***Aegithalos***[9], ***Leptopoecile***[9], ***Psaltriparus***[9]	**9** Aegithalidae	***Aegithalos, Leptopoecile, Psaltriparus*** *Psaltria*
		10 Phylloscopidae	***Phylloscopus, Seicercus***
		11 Acrocephalidae	***Nesillas, Acrocephalus, Iduna*** *(incl. Phragamaticola)****, Calamonastides*** *(Chloropeta)****, Hippolais***
		12 Locustellidae	***Dromaeocercus, Megalurus, Cincloramphus, Schoenicola, Eremiornis, Bradypterus, Locustella*** *Amphilais, Buettikoferella, Megalurulus, Chaetornis, Elaphrornis*
		13 Donacobiidae	***Donacobius***
		14 Bernieridae	***Oxylabes, Bernieria, Hartertula, Thamnornis, Xanthomixis, Crossleyia,*** *Cryptosylvicola, Randia*
Alaudidae	***Mirafra***[3], ***Ammomanes***[3], ***Alauda***[3] *Heteromirafra, Certhilauda, Pinarocorys, Chersomanes, Alaemon, Ramphocoris, Melanocorypha, Calandrella, Spizocorys, Eremalauda, Pseudalaemon, Galerida, Lullula, Eremopterix, Eremophila*		
Cisticolidae	***Cisticola***[15], ***Scotocerca***[8], ***Prinia***[15], ***Spiloptila***[15], ***Apalis***[15], ***Hypergerus***[15], ***Camaroptera***[15], ***Calamonastes***[15] *Incana, Rhopophilus, Schistolais, Phragmacia, Oreophilais, Heliolais, Urolais, Drymocichla, Phyllolais, Eminia, Euryptila*	**15** Cisticolidae	***Cisticola, Prinia, Spiloptila, Apalis, Hypergerus, Camaroptera, Calamonastes, Bathmocercus, Orthotomus, Artisornis, Eremomela*** *Neomixis, Incana, Schistolais, Phragmacia, Oreophilais, Heliolais, Urolais, Oreolais, Drymocichla, Phyllolais, Urorhipis, Malcorus, Eminia, Euryptila, Bathmocercus, Scepomycter, Poliolais*
Genera incertae sedis	***Orthotomus***[15], ***Artisornis***[15], *Poliolais, Neomixis*		
Pycnonotidae	***Pycnonotus***[4], ***Andropadus***[4], ***Chlorocichla***[4],		

Sylvioidea	Dickinson 2003	Sylvioidea	IOC World Bird Names List 2.10
Pycnonotidae	***Phyllastrephus***[4], ***Ixos***[4] *Spizixos, Calyptocichla, Baeopogon, Ixonotus, Thescelocichla, Pyrrhurus, Bleda, Criniger, Acritillas, Setornis, Tricholestes, Iole, Thapsinillas, Microscelis, Hemixos, Hypsipetes, Cerasophila*		
Genera incertae sedis	***Nicator***[2], ***Erythrocercus***[18], *Neolestes, Elminia*		
Sylviidae	Megalurinae:***Megalurus***[12], ***Cincloramphus***[12], ***Eremiornis***[12], ***Schoenicola***[12], *Amphilais, Buettikoferella, Megalurulus, Chaetornis, Graminicola* Acrocephalinae:***Tesia***[8], ***Cettia***[8], ***Bradypterus***[12, 7], ***Dromaeocercus***[12], ***Nesillas***[11], ***Melocichla***[7], ***Sphenoeacus***[7], ***Locustella***[12], ***Phragamaticola***[11], ***Acrocephalus***[11], ***Iduna***[11], ***Hippolais***[11], ***Chloropeta***[11], *Hemitesia, Oligura, Urosphena, Elaphrornis, Bathmocercus*		
Genera incertae sedis	***Macrosphenus***[17], ***Hylia***[8], ***Oxylabes***[14], ***Bernieria***[14], ***Hartertula***[14], ***Thamnornis***[14], ***Xanthomixis***[14], ***Crossleyia***[14], *Hyliota, Amaurocichla, Cryptosylvicola, Randia*		
	Phylloscopinae: ***Phylloscopus***[10], ***Seicercus***[10], ***Abroscopus***[8], ***Eremomela***[15], ***Sylvietta***[7], *Tickellia, Graueria* Sylviinae: ***Sylvia***[19], *Parisoma*		
Timaliidae	***Pellorneum***[17], ***Illadopsis***[17], ***Pseudoalcippe***[19], ***Pnoepyga***[6], ***Stachyris***[16], ***Dumetia***[16], ***Chrysomma***[19], ***Chamaea***[19], ***Turdoides***[18], ***Garrluax***[18], ***Alcippe***[17, 19], ***Phyllanthus***[18], ***Yuhina***[20], ***Erpornis****, ***Panurus***[1], ***Paradoxornis***[19] *Leonardina, Trichastoma, Malacocincla, Malacopteron, Kakamega, Ptyrticus, Pomatorhinus, Xiphirhynchus, Jabouilleia, rimator, Ptilocichla, Kenopia, Napothera, Spelaeornis, Sphenocichla, Rhopocichla, Macronous, Micromacronus, Timalia, Moupinia, Babax, Liocichla, Leiothrix, Cutia, Pteruthius, Gampsorhynchus, Actinodura, Minla, Lioptilis, Kupeornis, Parophasma, Crocias, Heterophasia, Conostoma*	**16** Timaliidae	***Dumetia, Stachyris*** *Pomatorhinus, Spelaeornis, Sphenocichla, Stachyridopsis, Rhopocichla, Macronus, Micromacronus, Timalia*
		17 Pellorneidae	***Illadopsis, Pellorneum*** *Alcippe, Napothera, Gampsorhynchus, Ptyrticus, Jabouilleia, Rimator, Malacocincla, Trichastoma, Leonardina, Robsonius, Kenopia, Graminicola*
		18 Leiothrichidae	***Phyllanthus, Turdoides, Trochalopteron*** ***(****Garrulax****)*** *Kupeornis, Babax, Garrulax, Cutia, Minla, Liocichla, Actinodura, Leiothrix, Crocias, Heterophasia*
		19 Sylviidae	***Pseudoalcippe, Sylvia, Lioparus*** ***(****Alcippe****), Chrysomma, Chamaea, Sinosuthora*** ***(****Paradoxornis****)*** *Myzornis, Parophasma, Horizorhinus, Lioptilus, Moupinia, Fulvetta, Rhopophilus,*

Sylvioidea	Dickinson 2003	Sylvioidea	IOC World Bird Names List 2.10
		19 Sylviidae	*Conostoma, Cholornis, Suthora, Neosuthora, Chleuasicus, Psittiparus, Paradoxornis*
Genera incertae sedis	*Chaetops*, Myzornis, Malia, Horizorhinus, Modulatrix*		
Zosteropidae	***Zosterops***[20], *Woodfordia, Rukia, Cleptornis, Apalopteron, Tephrozosterops, Madanga, Lophozosterops, Oculocincta, Heleia, Chlorocharis, Megazosterops, Hypocryptadius, Speirops*	**20** Zosteropidae	***Yuhina, Zosterops*** *Zosterornis, Megazosterops, Apalopteron, Cleptornis, Rukia, Sterrhoptilus, Tephrozosterops, Madanga, Lophozosterops, Heleia, Oculocincta, Woodfordia, Chlorocharis, Speirops*

2.2 Introduction to the families of Sylvioidea

In the following section a short introduction to all families is given. They are taken from Sibley and Ahlquist (1990), de Juana et al. (2004), Turner (2004), Bauer et al. (2005), Fishpool and Tobias (2005), Ryan (2006), Bairlein (2006), Collar and Robson (2007), Robson (2007), Harrap (2008), van Balen (2008), and Svensson et al. (2009). Number of genera etc. are taken from the IOC World Bird Names List (Gill and Donsker 2011). Families are listed and numbered in the same order as Table 1.

1) **Panuridae** (Bearded Reedling, Fig. 2): 1 genus, 1 species, 3 ssp.

Figure 2: *Panurus biarmicus*, 14-15 cm © Micha Luhn

The monotypic genus *Panurus* has three subspecies, which occur from Western Europe to NE China. The Bearded Reedling is a colonial breeder in large reed-beds at lowland lakes and swamp margins. *Panurus* was placed in Paradoxornithidae (parrotbills) before and was thus long regarded as the only European member of Timaliidae. Sexes coloured differently.

2) **Nicatoridae** (Nicators, Fig. 3): 1 genus, 3 species, 3 ssp.

Nicatoridae, another monotypic family, contains three species, and was placed in different groups before elevating it to its own family within Sylvioidea. First, it was regarded as a shrike, because of its coloration and bill morphology. Nicator was then moved to bulbuls (Pycnonotidae) as it shares some characteristics with them, like fluffy plumage, but this affiliation was never certain. The members of this family occur only in Africa.

Figure 3: *Nicator chloris*, 20-23cm, © del Hoyo et al. 2005

3) **Alaudidae** (Larks, Fig. 4): 20 genera, 97 species, 414 ssp.

Figure 4: *Alauda arvensis*, 16-18 cm, © Silke Fregin

Larks are a well defined group, also on morphological grounds. They occur mostly in Africa and Eurasia, a few species spread to Australia and America. This group can be delimited by two characteristic morphological features: 1) the back of the tarsus is covered with scales, whereas other oscines have smooth tarsi. 2) despite their ability to perform elaborate songs, Larks have, a relatively simple syrinx with only five instead of six to eight syrinx muscles and a rudimentary pessulus, a normally ossified structure at the bronchial junction. Because of these character states was the family of the larks placed at the basis of the oscines before they were moved to Sylvioidea.

4) **Pycnonotidae** (Bulbuls, Fig. 5): 27 genera, 49 species, 350 ssp.

One of the few families which have been relatively stable over the years. Bulbuls posses a long tail and short tarsi, and have soft and fluffy plumage. A characteristic feature is the ossified connective tissue in the nostril of the lightly hooked bill. Their main distribution is African and Oriental.

Figure 5: *Pycnonotus barbatus*, 15-20 cm, © Paul Vinke

5) **Hirundinidae** (Swallows and Martins, Fig. 6): 19 genera, 88 species, 206 ssp.

Figure 6: *Hirundo rustica*, 17-21 cm, © Silke Fregin

Members of the swallows and martins are very similar to each other because of their adaptation to aerial foraging on insects. As a rough distinction, "swallows" have more forked tails while "martins" are those with a more square tail. Due to their foraging behaviour, they are similar to swifts in appearance. Hirundinidae is the only family within Sylvioidea, where more than half of the species occur in the New World.

6) **Pnoepygidae** (Wren-babblers, Fig. 7): 1 genus, 4 species, 11 ssp.

Pnoepygidae are another monotypic family, split off the timaliid group relative recently (Gelang et al. 2009), based on a combination of genetic, morphological, and ecological characteristics. Among other characteristics, a nearly absent tail, cryptically patterned plumage with two colour morphs and their non-social behaviour delimits Pnoepygidae from timaliid species.

Figure 7: *Pnoepyga pusilla*, 7.5-9 cm, © del Hoyo et al. 2007

7) **Macrosphenidae** (Crombecs and African Warblers, Fig. 8): 6 genera, 18 species, 51 ssp.

Figure 8: *Sylvietta whytii*, 9 cm, © Paul Vinke

Macrosphenidae consists of African species of the former Sylviidae family. They forage for insects mostly on the ground or low in the vegetation from moist habitats near rivers to wooded savannas. Size ranges from ca. 10 cm in Crombecs to 20 cm in *Melocichla* and *Sphenoeacus*.

8) **Cettiidae** (Cettia Bush Warblers, Fig. 9): 12 genera, 39 species, 116 ssp.

Cettiidae possess only 10 tail feathers (*Tesia* even only eight), whereas most other sylvioid species have 12 rectrices. Many birds of this family are better distinguished by voice than appearance. Most members were in Sylviidae before Cettiidae were established by Alström et al. (2006a). Many occur in the Oriental or Australian region, with only one species, *Cettia cetti*, in the South of Central Europe.

Figure 9: *Cettia cetti*, 13-14 cm, © Silke Fregin

9) **Aegithalidae** (Long-tailed Tits, Bushtits, Fig. 10): 4 genera, 13 species, 49 ssp.

Figure 10: *Aegithalos caudatus*, 13-15 cm © Silke Fregin

Aegithalidae are more or less long-tailed insect eaters with short wings, which build a domed nest from moss, covered with cobwebs, hair, and lichen. Most members have an oriental distribution, one species occurs in North-America. The most widespread taxon is *Aegithalos caudatus*, breeding from Western Europe to China.

10) **Phylloscopidae** (Leaf-Warblers, Fig. 11): 2 genera, 77 species, 175 ssp.

Leaf-warblers are greenish, brown above and white or yellowish below, small birds with slender bills, and a pale supercilium. Often difficult to distinguish by morphology, songs, calls and molecular data are important for identifying these species. They forage in the foliage, but build their nests near or on the ground. Most species occur in the Oriental or Asian region, with only few species in Western Europe.

Figure 11: *Phylloscopus canariensis*, 11-12.5 cm, © Silke Fregin

11)**Acrocephalidae** (Reed-Warblers, Fig. 12): 5 genera, 55 species, 119 ssp.

Figure 12: *Acrocephalus schoenobaenus*, 11.5-13 cm, © Silke Fregin

Reed-warblers have brownish or greyish upperparts, some are tinged greenish with bright yellow underparts. Most occur in wet habitats such as reed-beds or marshland, some of them live in drier habitats (*Iduna* and *Nesillas*). Some species have a high degree of mimicry in their songs. They are distributed from Western Europe to Polynesia.

12) **Locustellidae** (Grass Birds, Fig. 13): 12 genera, 55 species, 127 ssp.

Locustellidae were renamed from Megaluridae (Alström et al. 2011). They are similar to reed-warblers with non-descript brownish upperparts; their songs are often simple, but important for identification and sometimes insect-like. They occur from Europe to China, in Africa and in the Australian region.

Figure 13: *Locustella naevia*, 12.5-13.5 cm, © Paul Vinke

13) **Donacobiidae** (Donacobius, Fig. 14): 1 genus, 1 species, 4 ssp.

Figure 14: *Donacobius atricapilla*, 21.5-22 cm, © del Hoyo et al. 2005

Donacobiidae is a monotypic family from South America. Before moving to Sylvioidea, it was placed in Troglodytidae (wrens) based on social behaviour and was also considered to be a member of Mimidae (e.g. mockingbirds) based on morphological characteristics, e.g. short rounded wings, long tail, heavy feet and legs. Only DNA studies showed that *Donacobius* belongs to Sylvioidea.

14) **Bernieridae** (Malagasy Warblers, Fig. 15): 12 genera, 11 species, 15 ssp.

Originally, the members of Bernieridae were distributed over three different groups, bulbuls, babblers and old world warblers Cibois et al. (1999, 2001). All species of this family are endemic to Madagascar. The family has been diagnosed only based on genetic data (Cibois et al. 2010), as they exhibit a great variety of morphological characteristics.

Figure 15: *Crossleyia xanthophrys*, 15 cm, © del Hoyo et al. 2006

15) **Cisticolidae** (Cisticolas, Fig. 16): 27 genera, 159 species, 534 ssp.

Cisticolidae is a largely African radiation, with some species occurring in the Oriental region and only one species also in South-West Europe, *Cisticola juncidis*. Cisticolas have mostly brown, grey or olive

Figure 16: *Cisticola juncidis*, 10-11cm, © Silke Fregin

upperparts, occur in open grassland or scrubs, and are difficult to distinguish in the field. The family is difficult to delimit by morphological characters and was defined through DNA data (Sibley and Ahlquist 1990, and citations therein).

16) **Timaliidae** (Babblers and Scimitar Babblers, Fig. 17): 10 genera, 56 species, 195 ssp.

Figure 17: *Timalia pileata*, 15.5 -17 cm, © del Hoyo et al. 2007

Timaliidae have often been called the "dustbin" of systematics, as many birds whose affiliations were not clear were placed in Timaliidae. According to the current understanding (Gelang et al. 2009, Gill and Donsker 2011), Timaliidae are only a subset of former Timaliidae. They are very sociable birds, but a rather diverse group with respect to size, habit and vocalization. Timaliidae have a restricted distribution in Asia, from India/SW China to Borneo. They live mostly in thickets or forest undergrowth. This family includes the smallest babbler genus, *Micromacronus*, which weighs only 5.5 g and is 7-8 cm long.

17) **Pellorneidae** (Ground Babblers, Fig. 18): 15 genera, 70 species, 219 ssp.

Figure 18: *Pellorneum capistratum*, 16-17 cm, © del Hoyo et al. 2007

Pellorneidae was split off the former Timaliidae, which have also an Asian distribution from the Himalaya region to Borneo and Sumatra; two genera also occur in Africa. They can be found in forests, forest undergrowth and thickets.

18) **Leiothrichidae** (Laughingthrushes, Fig. 19): 13 genera, 133 species, 328 ssp.

Figure 19: *Turdoides leucopygia*, 25-27 cm, © Paul Vinke

Leiothrichidae was also split off the former Timaliidae; they occur from the Arabian peninsula to Indonesia and some species live also in Africa. They inhabit a wide range of habitats from arid scrubs, open areas to forests and forest undergrowth. This family contains the largest babbler, *Garrulax pectoralis*, with a weight of 170 g and a length of 34.5 cm.

19) **Sylviidae** (Sylviid Babblers, Fig. 20): 20 genera, 70 species, 185 ssp.

Figure 20: *Sylvia curruca,* 11.5-13.5 cm, © Silke Fregin

Sylviidae are medium-sized warblers with robust bill and legs. They have an Eurasian and African distribution and one species occurs in North America. Their habitats are mostly drier encompassing scrubs, thickets, grassland and bamboo.

20) **Zosteropidae** (White-Eyes, Fig. 21): 17 genera, 125 species, 305 ssp.

The name of the family is derived from the very small white feathers around the eyes of many species. Zosteropidae are absent from Central Europe and have a mostly Asian or Australopacific distribution; some occur in Africa, where they live mostly in forests. Their 10th primary, the outermost wing feather, is reduced or absent. Some genera, *Zosterops* and *Yuhina*, are adapted to nectar-feeding; *Zosterops* was therefore formerly associated with other nectar-feeding birds like honeyeaters and sunbirds.

Figure 21: *Zosterops poliogastrus,* 11.5-12 cm, © Paul Vinke

2.3 The Goals of the thesis

In my doctoral thesis I investigated several aspects of the systematics of the avian superfamily Sylvioidea based on molecular data. As outlined above, the group is very challenging and many studies were carried out to clarify especially species affiliations within single families. Only some analyses aimed at the relationships between families (Alström et al. 2006a, Fuchs et al. 2006, Johansson et al. 2008, Irestedt et al 2010). My analyses focused at three different levels:

1) On the family level I analysed the relationships between the families within the superfamily Sylvioidea (Fregin et al. submitted).
2) On the genus level I aimed at clarifying the relationships within two families, viz. the Acrocephalidae and Locustellidae (Fregin et al. 2009, Alström et al. 2011b).
3) And on the species level I studied the applicability and problems of barcoding based on one sylvioid family, the Acrocephalidae (Fregin et al. 2012).

3 Introduction to individual studies

3.1 New insights into family relationships within the avian superfamily Sylvioidea (Passeriformes) based on seven molecular markers

The first step in reorganizing Sylvioidea was conducted by Alström et al. (2006a). They identified ten well supported major clades, which they proposed be recognized at the family level, namely: Alaudidae, Pycnonotidae, Hirundinidae, Cettiidae, Aegithalidae, Phylloscopidae, Timaliidae, Acrocephalidae, Megaluridae, and Cisticolidae. Their findings were supported by Johansson et al. (2008), and complemented with some additional clades, which were found in need of further research previously, such as the genus *Nicator*, the "malagasy warblers" now called Bernieridae (Cibois et al. 2010) and the supposed flycatcher *Erythrocercus*. While the next relatives to Bernieridae were quite well established, the positions of *Nicator* and *Erythrocercus* were less clear. Another important step was the reclassification of the sylviid/timaliid group (Gelang et al. 2009). Gelang et al. (2009) re-established the family Sylviidae, which was temporarily lost, as it was merged with Timaliidae. The type genus of Sylviidae Sylvia (Leach, 1820) was shown to be nested within the large Timaliidae (Vigors and Horsfield, 1827) assemblage. Based on these results, it was suggested to suppress Sylviidae (Cibois 2003a, Alström et al. 2006a), following the principle of stability (Cibois 2003b). But this would have implied a taxonomic conflict, if the superfamily Sylvioidea would have lost its name-giving family.
Despite these analysis which involved partly large datasets (Alström et al. 2006a: 1800bp (basepairs) and 83 species, Johansson et al. 2008: ~2300bp and 45 species or 7400bp and 14 species) basal relationships within Sylvioidea remained mostly unclear. We inferred relationships within Sylvioidea based on 6300bp and 79 taxa, including all known families at this point.

3.2 Multi-locus phylogeny of the family Acrocephalidae (Aves: Passeriformes) – The traditional taxonomy overthrown

The Reed-Warbler family Acrocephalidae sensu Alström et al. (2006a) and Johansson et al. (2008) consists of four genera *Acrocephalus*, *Hippolais*, *Chloropeta*, and *Nesillas*. Relationships among Reed-Warblers have long been discussed based on morphology and oology (Voous 1975, Meise 1976, Schönwetter 1979, Wolters 1982, Watson et al. 1986, Haffer 1991, Cramp and Perrins 1993, Dickinson 2003). But due to their morphological similarity (e.g. Kennerley and Leader 1992, Cramp and Perrins 1993, Baierlein 2006) and convergent adaptations to similar habitats (Leisler 1980, Leisler et al. 1989, Haffer 1991, Baierlein 2006) many species are difficult to distinguish. Also analysis based on a single

mitochondrial gene (Leisler et al. 1997, Helbig and Seibold 1999) were not sufficient to resolve the relationships between the different genera, as Acrocephalidae have probably evolved in a rapid radiation (e.g. Sibley and Ahlquist 1990, Parkin et al. 2004).

This study is an extension of Leisler et al (1997) and Helbig and Seibold (1999) in respect of taxon sampling and number of DNA sequences. Both used incomplete mitochondrial cytochrome b sequences. We included all four genera of Acrocephalidae for the first time in one analysis, complemented the cytochrome b sequences to full length and expanded the data set with three nuclear markers, in total ~2900bp.

3.3 Multilocus analysis of a taxonomically densely sampled dataset reveal extensive non-monophyly in the avian family Locustellidae

The family Megaluridae sensu Alström et al. (2006a) was derived from the subfamily Megalurinae sensu Sibley and Monroe (1990) which contained the genera *Megalurus*, *Cincloramphus*, *Eremiornis*, *Amphilais*, *Megalurulus*, *Buettikoferella*, *Chaetornis*, *Graminicola* and *Schoenicola*. But the genera found to be included in Megaluridae by Alström et al. (2006a) and later by Johansson et al. (2008) were *Megalurus*, *Bradypterus*, *Locustella* and *Dromaeocercus*, which were partly included in Acrocephalinae (Sibley and Monroe 1990, see also Table 1). A previous study showed that also *Cincloramphus* and *Schoenicola* belong to this clade (Beresford et al. 2005). Megaluridae had to be renamed to Locustellidae, as Locustellinae Bonaparte, 1854, has priority over Megalurinae Blyth, 1875 (Bock, 1994: p 152).

Drovetski et al. (2004), based on the mitochondrial ND2 gene (1041bp, NADH-ubiquinone oxidoreductase chain 2) and 19 species, found the Asian and African *Bradypterus* species in separate clades, and the Asian clade was nested within *Locustella*. Also one of the two included *Megalurus* species (*Megalurus pryeri*) was recovered within *Locustella*. Thus, indicating the need of a more extensive research. We reanalysed the family Locustellidae with one mitochondrial and four nuclear markers, in total ~3300 aligned bp and 37 species from seven genera.

3.4 Pitfalls in comparisons of genetic distances: A case study of the avian family Acrocephalidae

Genetic distances are a frequently used tool for identification and assessing species status of closely related taxa (e.g. Wesson et al. 1993, Hung et al. 1999, Burbrink et al. 2000, Bradley and Baker 2001, Cagnon et al. 2004, Parkin et al. 2004, Olsson et al. 2005, Newman et al. 2012). In this way, genetic

distances are often frequently compared between different studies, even if different genetic markers are involved (e.g. Baker et al. 2003, Helbig et al. 1995, Johnson and Cicero 2004, Zhang et al. 2007). Not only different loci are compared, but also genetic distances, which were not estimated based on the best fitting substitution model. Mostly, under-corrected or even uncorrected distances are used (e.g. Benzoni et al. 2010, Loader et al. 2010, Palma et al. 2010). This could result in an underestimation of the actual genetic distance between a species-pair (Arbogast et al. 2002). Model selection is considered very important in phylogenetic analysis, because branch lengths and thus probably the tree topology are influenced by the substitution model (Yang et al. 1994, Posada and Crandall 2001, Johnson and Omland 2003, Lemmon and Moriarty 2004, Posada and Buckley 2004, Posada 2008). Barcoding is based on the existence of a "barcoding gap", which is supposed to separate mean intraspecific genetic divergence from mean genetic divergence between species. It is expected to be 10 times higher than the latter (Hebert et al. 2004), and a maximum divergence of 2% is considered to be threshold of delimiting species.

The purpose of this study was to examine how genetic distances between sequences of the mitochondrial gene cytochrome b are affected by different methods of calculation:

1) uncorrected p-distances vs. distances derived from the best-fit model;
2) comparison of different parts of one locus under the same correction model;
3) after "complete deletion" vs. "pairwise deletion"; often, sequences are of different length or contain a certain amount of unknown nucleotides "N"; complete deletion neglects the position of an unknown nucleotide over the whole alignment, whereas pairwise deletion just removes this position for the particular sequence-pair under comparison;

Finally, it was investigated, how these differences affect the existence of a barcoding gap in Acrocephalidae.

4 Materials and methods

4.1 DNA extraction, amplification, sequencing and assembly

DNA was extracted with a salting out procedure according to Miller et al. (1988) with slight modifications or using a QIAamp® DNA MiniKit (50) following the manufacturer's protocol. The following genetic markers were used: cytochrome b (MB–CYB, ≤1143bp), fibrinogen beta chain intron 5 (FGB, ≤570bp), glyceraldehyd–3–phosphat dehydrogenase (GAPDH, ≤395bp), lactate-dehydrogenase B intron 3 (LDHB, ≤520bp), myoglobin intron 2 (MB, ≤705bp), ornithine-decarboxylase exon 6–8, intron 7 (ODC1, ≤730bp), recombination activation gene 1 (RAG1, ≤1936bp). The fibrinogen beta chain intron 5 was retrieved from GenBank.

Primers for amplification and sequencing are given in Table 2. The MB–CYB gene was, whenever possible, amplified including flanking parts to reduce the risk of amplifying nuclear copies (numts, nuclear mitochondrial DNA) (Sorenson and Quinn 1998). Sequences were therefore checked for unexpected stop–codons, but none were detected. PCR products were cleaned with ExoSap IT, which consists of Exonuclease I and Shrimp Alkaline Phosphatase for removing excess nucleotides and single–stranded DNA including primers. Products from cycle sequencing were cleaned with DyeEx 96Plate from Quiagen, only when the ABI sequencer was used. Sequencing was done on a LiCor DNA Sequencer Long READIR 4200 or on an ABI 3130xl Genetic Analyzer. Sequences were assembled manually in BioEdit (Hall 1999) or with the Staden Package (Bonfield et al. 1995).

4.2 Phylogenetic analysis

Sequences were aligned by eye in BioEdit (Hall 1999), with the online program T–Coffee (Notredame et al. 2000) or with MAFFT (Katho et al. 2002) with eventually subsequent manual adjustments. Phylogenetic analyses were performed by Bayesian inference (BI) using MrBayes 3.1 (Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003), maximum likelihood (ML) with Treefinder version October 2008 (Jobb et al. 2004, Jobb 2008) or GARLI–PART 0.97 (Zwickl 2006), and maximum parsimony (MP) in PAUP* (Swofford 2003). MrModeltest (Nylander 2004) was used in conjunction with PAUP* to estimate appropriate nucleotide substitution models for implementation in MrBayes, based on the Akaike Information Criterion (AIC; Akaike 1973) and AICc for small sample sizes (Sugiura 1978, Hurvich and Tsai 1989). For the ML analyses in Treefinder its internal model proposer was used to estimate nucleotide substitution models based on the AIC. When GARLI–PART was used, best fit nucleotide substitution models were estimated with jModelTest (Posada 2008, Guindon and Gascuel 2003), with the same criteria as for MrModeltest. Convergence of parameters in Bayesian

analysis was monitored using the program Tracer v. 1.4 (Drummond and Rambaut, 2007). In addition, the average standard deviation of split frequencies across both independent runs was controlled to ascertain convergence of the runs suggested by a value below 0.01. Every 100th of the gene-dependent total number of generations was sampled. All genetic markers were analysed separately and combined. Number of non-parametric bootstrap replicates varied between individual studies. Posterior probabilities ≥ 0.95 and bootstrap values ≥ 85% were regarded as high support following Erixon et al. (2003).

Table 2: Primers used in amplification and sequencing. If not indicated otherwise, primers were newly designed. Forward primers are given in bold.

gene	name	sequence	reference
MB–CYB	mtB–Acro	5'–CCT CAG AAT GAT AT(CT) TGT CCT CA–3'	
	mtB–NP	5'–CCT CAG AAG GAT AT(CT) TG(GT) CCT CA–3'	Helbig et al. 2005
	mtD	**5'–GA(CT) AAA ATC CCA TT(CT) CA(CT)–3'**	Helbig et al. 2005
	mtD–Syl	**5'–GAC TG(CT) GA(CT) AAA AT(CT)CCA TTC CA–3'**	
	mtE–Syl	5'–GGG TT(AG) GC(GT) GG(GCT) GTG AA(AG) TTT TC–3'	
	mtE	5'–GGG TTT GCT GGG GT(AG) AA(AG) TTT TC–3'	Helbig et al. 2005
	mtE–Acro	5'–GGG TTG GC(GT) GG(CT) GTG AA(AG) TTT TC–3'	
	mtF–21	5'–GGT TTA CAA GAC CAA TGT TTT–3'	
	mtF–NP	5'–GG(CT) TTA CAA GAC CAA TGT TT–3'	
	mtF–24	5'–TTT GGT TTA CAA GAC CAA TGT TTT–3'	
	mtL–Syl	5'–GCG TA(GT) GC(AG) AAT A(AG)G AAG TA–3'	
	ND5–23	**5'–GG(AG) CTA (AC)T(CT) AAA ACC TAC (CT)TA GG–3'**	
	tRNA–Leu	**5'–A(CT)C TTG GTG CAA ATC CAA GT–3'**	Helbig et al. 2005
	H15915	5'–AAC TGC AGT CAT CTC CGG TTT ACA AGA C–3'	Edwards and Wilson 1990
	L14841	**5'–CCA TCC AAC ATC TCA GCA TGA TGA AA–3'**	Kocher et al. 1989
GAPDH	**13b**	**5'–AAG TCC ACA ACA CGG TTG CTG TA–3'**	Fjeldså et al. 2003
	14b	5'–TCC ACC TTT GAT GCG GGT GCT GGC AT–3'	Fjeldså et al. 2003
	intLI	**5'–GAA CGA CCA TTT TGT CAA GCT GGT T–3'**	Fjeldså et al. 2003
	F1	**5'–CAA (CT)GG TGA CAG CCA TTC CTC CAC–3'**	
	R1	5'–GTC CAC AAC ACG GTT GCT GTA TCC–3'	
LDHB	**b1**	**5'–GGA AGA CAA ACT AAA AGG AGA AAT GAT GGA–3'**	Friesen et al. 1999
	b4	5'–GGG CTG TAT TT(AGCT) AC(AG) ATC TGA GG–3'	Helbig et al. 2005
	P1	**5'–ACTA AAA GG(AG) GAA ATG ATG GA–3'**	Helbig et al. 2005
	P5	**5'–GCT TGC TCT GGT TGA (CT)GT TAT GG–3'**	
	P6	5'–CAC ATT CCT CTG CAC (CT)AG GTT GAG–3'	
	P6	5'–ATT CCT CTG CAC (CT)AG GTT GAG–3'	
MB	**Myo2**	**5'–GCC ACC AAG CAC AAG ATC CC–3'**	Slade et al. 1993
	Myo3f	5'–GCC ACC AAG CAC AAG ATC CC–3'	Heslewood et al. 1998
ODC1	**6**	**5'–GAC TCC AAA GCA GTT TGT CGT CTC AGT GT–3'**	Allen and Omland 2003
	8r	5'–TCT TCA GAG CCA GGG AAG CCA CCA CCA AT–3'	Allen and Omland 2003
	intF2	**5'–CAC TTA AGA CTA GCA GGC TTC TTC TGG A–3'**	Irestedt et al. 2006
	intR3	5'–CAA ACA CAC AGC GGG CAT CAG A–3'	Irestedt et al. 2006
	intF1	**5'–ATG CCC GCT GTG TGT TTG–3'**	Irestedt et al. 2006
	intR4	5'–CAT ATT GAA GCC AAG TTC AGC CTA–3'	Irestedt et al. 2006

gene	name	sequence	reference
RAG1	**F2–Pass**	**5'–TAC CCA (GC)TA GAT TTC ATT AAA TC–3'**	
	FL1	5'–GGG AAG CAA GGA TAC CAG CAG G–3'	
	FL1–Pass	5'–AGG (AT)GG GGC AAT AGC TGC CCA TAA–3'	
	S1	5'–TTC AGG AAG GAT TTC ACT GG–3'	
	S3–Pass	5'–GGA TAG T(AG)T C(AG)A CAT CCC CTC GCA–3'	
	S4–Pass	**5'–TTG (CT)GA TCA TAT TTT GGC AGA TCC–3'**	
	R3000	5'–TGA GCA TTC ATG AAC TTC TGG AG–3'	
	2I	5'–GAG GTA TAT AGC CAG TGA TGC TT–3'	Barker et al. 2002
	13C	**5'–TCT GAA TGG AAA TTC AAG CTC TT–3'**	Groth and Barrowclough 1999
	R13	**5'–TCT GAA TGG AAA TTC AAG CTG TT–3'**	Groth and Barrowclough 1999
	R15	**5'–TCG CTA AGG TTT TCA AGA TTG A–3'**	Groth and Barrowclough 1999
	R16	5'–GTT TGG GGA GTG GGG TTG CCA –3'	Groth and Barrowclough 1999
	R17	**5'–CCC TCC TGC TGG TAT TCC TTG CTT–3'**	Groth and Barrowclough 1999
	R18	5'–GAT GCT GCC TCG GTC GGC CAC CTT T –3'	Groth and Barrowclough 1999
	R19	**5'–GTC ACT GGG AGG CAG ATC TTC CA–3'**	Groth and Barrowclough 1999
	R20	5'–CCA TCT ATA ATT CCC ACT TCT GT–3'	Groth and Barrowclough 1999
	R21	**5'–GGA TCT TTG AGG AAG TAA AGC CCA A–3'**	Groth and Barrowclough 1999
	R22	5'–GAA TGT TCT CAG GAT GCC TCC CAT–3'	Groth and Barrowclough 1999
	R23	**5'–TAC AAG AAT CCT GAT GTG TCT AA –3'**	Groth and Barrowclough 1999
	R24	5'–GCC TCT ACT GTC TCT TTG GAC AT–3'	Groth and Barrowclough 1999

5 Results and discussions of individual studies

5.1 New insights into family relationships within the avian superfamily Sylvioidea (Passeriformes) based on seven molecular markers

Both inferences methods, Maximum likelihood (ML) and Bayesian inference (BI), resulted in identical trees regarding relationships within Sylvioidea. Although slightly different in the constellations of the outgroup taxa, both methods showed that Paridae, Remizidae and Stenostiridae clustered within the outgroup taxa. These families were sometimes still included in Sylvioidea (e.g. Sibley and Monroe 1990, Cracraft et al. 2004, Harshman 2006, Sangster et al. 2010). We thus recommended the exclusion of these families from Sylvioidea, as they were also treated as separate taxa before (e.g. Fuchs et al. 2006, Johansson et al. 2008, Treplin et al. 2008).

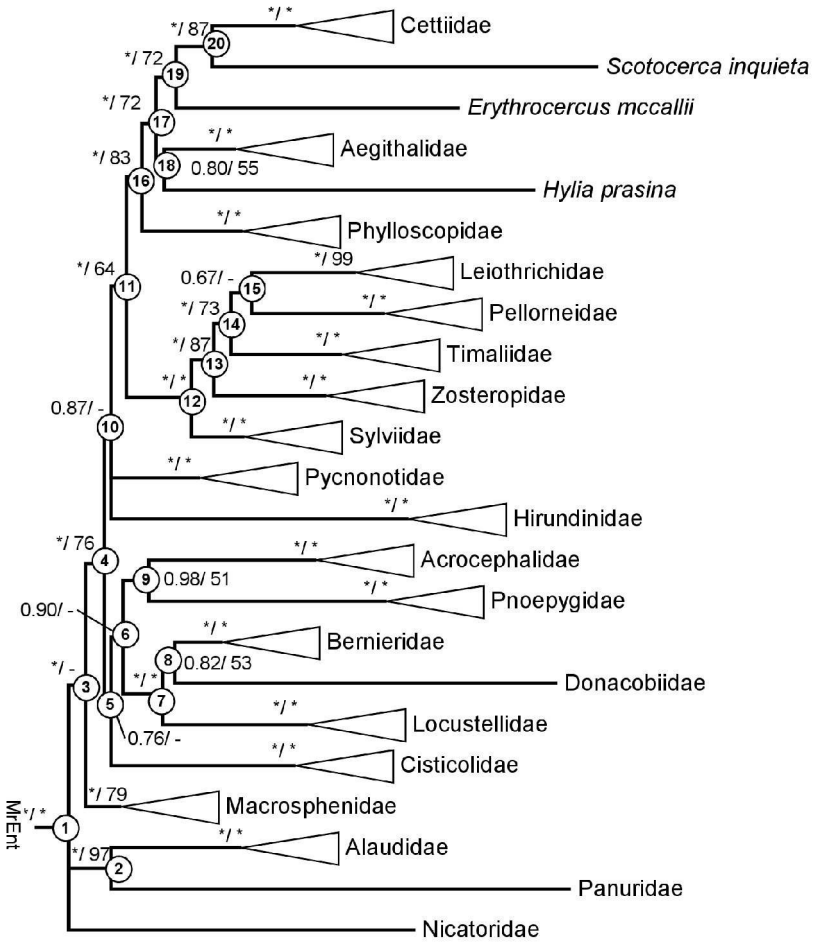

Figure 22: Phylogenetic tree of Sylvioidea based on the complete dataset (MT-CYB, FGB, GAPDH, LDHB, MB, ODC1, RAG1), analysed by Bayesian inference. Support values are given in the order posterior probability (PP) / maximum likelihood (ML) bootstrap; an asterisk indicates PP 1.00 or ML 100%, minus indicates no support. Outgroup has been pruned from the tree.

The superfamily Sylvioidea itself received very high support in all concatenated analysis. The close relationship of Alaudidae and Panuridae (node 2, Fig. 22) was very well supported, also by single locus analysis and a six basepair insertion in ODC1. *Panurus* was once considered to be the only European timaliid species (e.g. Sibley and Monroe 1990, Dickinson 2003). The position of Nicatoridae and Macrosphenidae and their relationship with Alaudidae/Panuridae is still not resolved, but it is getting somewhat clearer that these four families constitute basal splits within Sylvioidea. Macrosphenidae was sometimes found to be nested within a more derived position. This was mostly the case when only one or two genes were used for analyses (Alström et al. 2006a, Fuchs et al. 2006). We found Pnoepygidae, a newly established family by Gelang et al. (2009) as sister to Acrocephalidae, this was only well supported BI and was not found by Gelang et al. (2009) in their combined analysis. A very well supported grouping of several families were Locustellidae, Donacobiidae and Bernieridae,

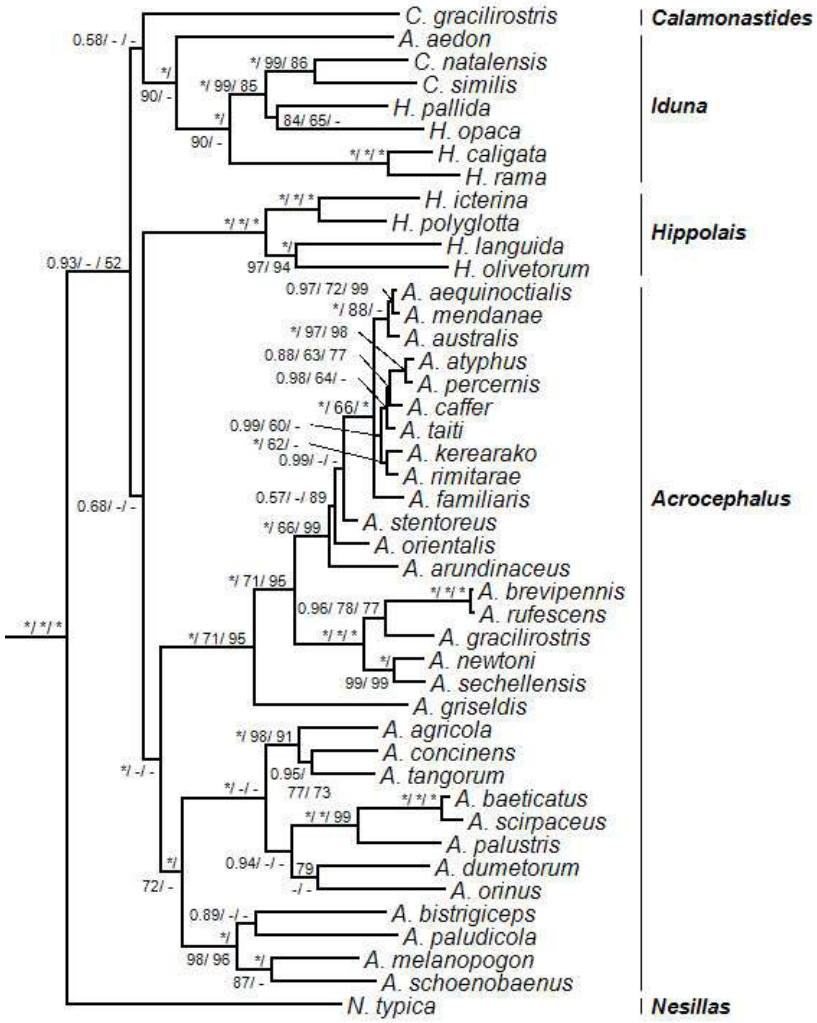

Figure 23: Phylogenetic tree of Acrocephalidae based on the complete concatenated dataset (MT-CYB, LDHB, MB, ODC1) analysed by Bayesian inference. Support values are given in the order posterior probability (PP) / maximum likelihood (ML) bootstrap / maximum parsimony bootstrap (MP); an asterisk indicates PP 1.00, ML or MP 100%, minus indicates no support. A.: *Acrocephalus*, C.: *Chloropeta*, H.: *Hippolais*, N.: *Nesillas*. Names on the right show revised taxonomy.

although the position of Donacobiidae is still not well supported. The relationships of Hirundinidae and Pycnonotidae to the remaining families in clade 10 were not resolved. The clustering of Sylviidae, Zosteropidae, Timaliidae, Leiothrichidae, and Pellorneidae was very well supported in both analyses. The relationships within this clade differed between the analyses of the complete data set and the analysis of the nuclear data set. In addition, single locus analyses resulted in different relationships, sometimes with high support. Thus, explaining the partly not well supported nodes between families within this clade.

The proposed affiliation of *Hylia* to Cettiidae by Gill and Donsker (2011), albeit tentatively, turned out to be better reversed, as this would render Cettiidae non-monophyletic in our analyses. *Hylia* has been placed in a family Hyliidae (Bates 1930) before, together with a presumptive sister genus *Pholidornis*. This sister relationship was found based on ND2 and 12S (Sefc et al. 2003), and anatomical details (Bates 1930). We therefore tentatively support the recognition of the family Hyliidae Bates, 1930 for *Hylia* and *Pholidornis*.

Scotocerca and *Erythrocercus* were identified as nearest relatives of Cettiidae (sensu Alström et al. 2006a). But based on their separation from Cettiidae by long branches and their morphological and ecological differences we propose to treat them as two monotypic families, Scotocercidae and Erythrocercidae.

5.2 Multi-locus phylogeny of the family Acrocephalidae (Aves: Passeriformes) – The traditional taxonomy overthrown

Neither Leisler et al. (1997), nor Helbig and Seibold (1999) could support the monophyly of the genera *Acrocephalus* and *Hippolais*, as the deeper nodes of the phylogeny were not resolved. The aim of this study was to address this question, with addition of three nuclear markers to the mitochondrial cytochrome b gene, ~2900bp in total.

In our analysis, neither *Acrocephalus*, nor *Hippolais* or *Chloropeta* were recovered as monophyletic (Fig. 23). One species of *Acrocephalus*, *A. aedon,* was sister to a clade consisting of four *Hippolais* species, which were referred to as *Iduna* by Helbig and Seibold (1999) and Leisler et al. (1997), and two *Chloropeta* species, *C. natalensis* and *C. similis*. This relationship was only resolved by the ODC1 gene, but additional evidence came from two indels in two independent genetic markers, 10bp insertion in ODC1 and 14bp deletion in LDHB. *A. aedon* is very different in appearance, being much larger and differently shaped than the other species in this clade. But its egg coloration is more similar to *Hippolais* and *C. natalensis*, than to other *Acrocephalus* species (Schönwetter 1979). In addition, its habitat preferences is also in congruence with the *Iduna* species (Alström et al. 2006b), preferring drier

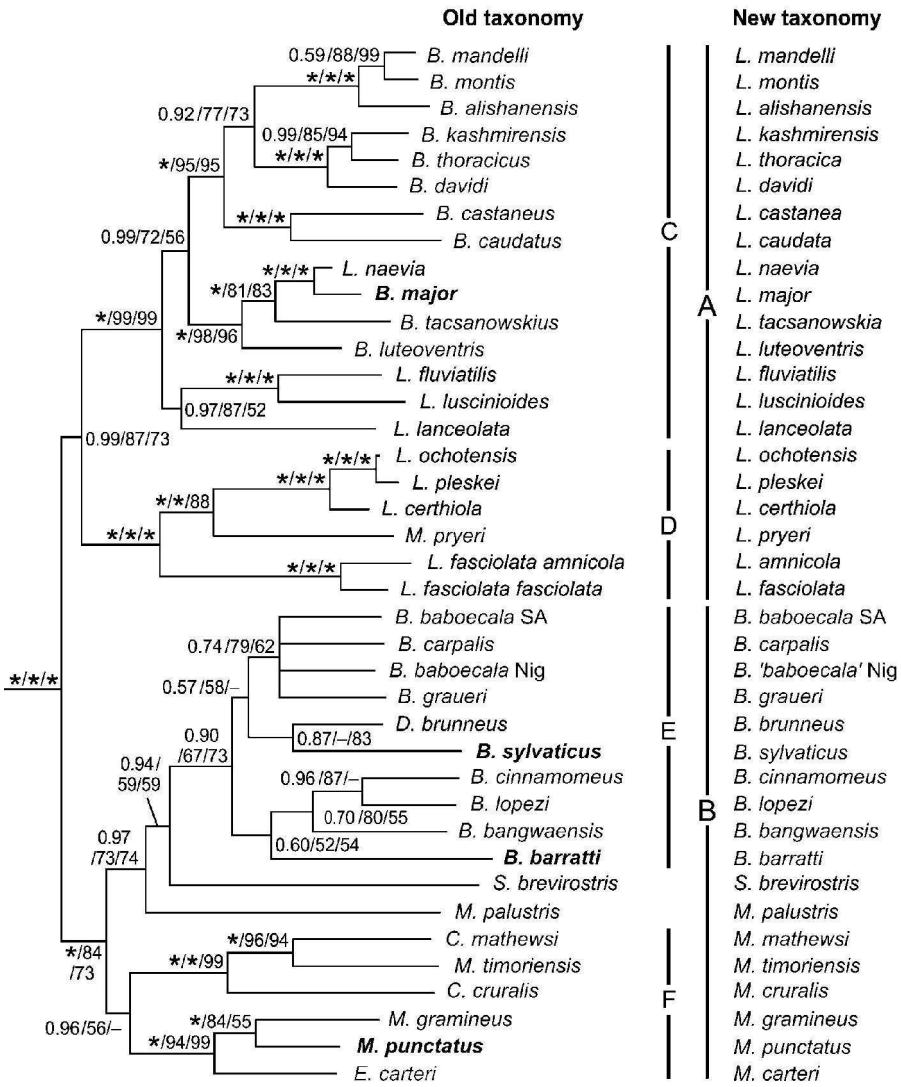

Figure 24: Phylogenetic tree of Locustellidae based on the complete concatenated dataset (MT-CYB, GAPDH, LDHB, MB, ODC1) analysed by Bayesian inference. Support values are given in the order posterior probability (PP) / maximum likelihood (ML) bootstrap / maximum parsimony bootstrap (MP); an asterisk indicates PP 1.00, ML or MP 100%, minus indicates no support. B.: *Bradypterus*, C.: *Cincloramphus*, D.: *Dromaeocercus*, E.: *Eremiornis*, L.: *Locustella*, M.; *Megalurus*, S.: *Schoenicola;* SA: South Africa, Nig: Nigeria. Names in bold refer to samples were only cytochrome b was available.

habitats with thickets.

Marova et al. (2005), came also to the conclusion that *A. aedon* should be separated from the remaining *Acrocephalus* species, based on nest construction and vocalisation, which is more like *Hippolais*.

Hippolais is clearly not monophyletic, which could also be shown by an approximately unbiased test. The species of *Hippolais* fell in two different clades, *Hippolais* s.s. and the *Hippolais* species which were referred to as *Iduna* by Leisler et al. (1997) and Helbig and Seibold (1999). These two clades did not cluster together and in addition two of the three *Chloropeta* species clustered in the *Iduna*-clade. *Chloropeta natalensis* and *C. similis* have not been included in the previous studies. Both fit well to the other species in this clade, in respect of size, habitat use and range of distribution. The position and

next relative of *Chloropeta gracilirostris* could not be resolved.

Due to these results we proposed the following revised taxonomy for the Reed-Warbler family Acrocephalidae: 1) All *Acrocephalus* species, with the exception of *A. aedon* remain in the genus *Acrocephalus*. 2) *Hippolais* should be restricted to *Hippolais* s.s. 3) *Iduna* should be the genus name for the remaining *Hippolais* species, *Chloropeta similis*, *Chloropeta natalensis*, and *Acrocephalus aedon*. 4) *Chloropeta gracilirostris* should be placed in a monotypic genus *Calamonastides*, as it did not show an affinity to either of the foregoing clades.

5.3 Multilocus analysis of a taxonomically densely sampled dataset reveal extensive non-monophyly in the avian family Locustellidae

The Locustellidae fell in two major clades (Fig. 24), clade A contained all included *Locustella* species, all Asian/Oriental species of *Bradypterus*, and one *Megalurus* species, *M. pryeri*. Clade B included the African *Bradypterus*, the monotypic Malagasy genus *Dromaeocercus*, the African *Schoenicola brevirostris*, the two Australian species of *Cincloramphus*, the monotypic Australian genus *Eremiornis* and four species of *Megalurus* (south Asia to Australasia). This division is supported by one indel each and is in accordance with biogeography, with one Palearctic/Asian/Oriental clade and one Afrotropical/Oriental to Australasian clade (Bairlein et al. 2006). Based on this topology we propose all species in clade A to be included in the genus *Locustella*. *Locustella naevia*, the type species, is nested within the African *Bradypterus* species. This clade is well supported in different analyses and also in accordance with biogeography. *Bradypterus* should tentatively be limited to clade E, which includes the type species *B. baboecala*. We excluded *Schoenicola* from this revision, as the second species of *Schoenicola* (*S. platyurus*), which has an Asian distribution, in contrast to the African *S. brevirostris*, and two missing African *Bradypterus* species were not included in thus analysis. We suggest *Megalurus* for clade F. Several samples which have been suggested to be closely related to *Megalurus* were lacking for this study. Thus, we refrain renaming the whole clade B until further research show how these missing samples are related to the taxa included in this study. In addition, the position of the type species of *Megalurus*, *M. palustris*, is somewhat unclear. Although it is mostly well supported in the complete concatenated analysis, its position is unresolved when using all nuclear genetic markers together, without the mitochondrial cytochrome b. Beresford et al. (2005), based on nuclear RAG1 and RAG2 sequences, found *M. palustris* clustering with *Cincloramphus* and not with *Schoenicola*/*B. barratti*/*B. baboecala*. It was not included in the study of Drovetski et al. (2004).

5.4 Pitfalls in comparisons of genetic distances: A case study of the avian family Acrocephalidae

The cytochrome b alignment comprised 879 nucleotides, including several uncertain positions (Ns). After applying complete deletion, only 718 nucleotides were effectively usable for the whole alignment. Using pairwise deletion, the compared sequence length varied between different pairwise comparisons. These variations were the basis for comparing the four different methods of distance calculation (see section 3.1).

Uncorrected p-distances vs. distances derived from the best-fit model

The best-fit substitution model for this alignment was determined as the Tamura–Nei model (TrN, Tamura and Nei 1993), assuming rate variation across sites according to a discrete gamma distribution with four rate categories (Γ, Yang et al. 1994). As expected, differences between p-distances and corrected distances were markedly higher between genera than within genera.

Testing the 2% threshold, which was found to be the limit for intraspecific differences by Ratnasingham and Hebert (2007), 76/65 (uncorrected/corrected) pairwise sequence comparisons had lower than 2% divergence. This involved 14 taxa currently treated as full species, mainly species from Pacific islands and two sister pairs *Acrocephalus brevipennis* / *Acrocephalus rufescens* and *Acrocephalus baeticatus*/*Acrocephalus scirpaceus*. Thus, in our cytochrome b dataset for the Acrocephalidae, we did not find a barcoding gap, as was recovered by Aliabadian et al. (2009) for cytochrome b in a wide range of birds. The reason for this result might be an inconsistent taxonomic treatment and/or recent speciation (Cibois et al. 2011). When applying the best-fit substitution model, less species pair comparisons fell under the 2% threshold indicating that some species would have been misidentified due to underestimated distances using uncorrected genetic distances. We therefore recommend that only genetic distances derived from the best-fit substitution model should be used.

Comparing sequences of different length/different parts of the same locus

After applying complete deletion, the cytochrome b alignment was reduced from 879 to 719 nucleotides, which also reduced the genetic divergence between species in most cases. Two sequences with no missing/unknown nucleotides had a sequence divergence of 0.034 (corrected) with 879 nucleotides, and 0.020 with 719 nucleotides. This highlights the importance of comparing identical lengths of the same parts of a genetic marker. However, some pairwise comparisons were not affected by the choice of method. The same principles apply for comparing different parts of cytochrome b. In general, the first half of cytochrome b produced higher genetic distances, than the second half (U-test: $p < 0.001$) demonstrating the importance to use identical/homologous parts of a sequence for

comparison. Although the first half showed higher genetic distances in general, there were also cases were the reverse was the case. In one case in *Acrocephalus melanopogon mimicus* intraspecific distance increased from 0.010 to 0.027 in the second half. A change that would also influence results when applying the 2% barcoding threshold for species delimitation. It was also demonstrated for CO1 and CO2 that different parts of the same gene vary substantially (Roe and Sperling 2007). Hebert et al. (2003b) also compared both halves of CO1 and found that the halves differed by 2.3%, with a standard deviation of 6.2%. Strangely, they concluded that this is similar enough and therefore not necessitating the specification of the source region within the gene.

Considering the numerous pitfalls for calculating genetic distances defined in this study and by foregoing studies (Roe and Sperling 2007, Palma et al. 2010, Schenk and Hufford 2010) it is recommended to use integrative approaches in taxonomy (e.g. Elias et al. 2007, Gomez et al. 2007, Haase et al. 2007, Ratnasingham and Hebert 2007, Alström et al. 2008, Capa et al. 2010, Ekrem et al. 2010, Markolf et al. 2011), instead of relying solely on species discrimination based on a certain threshold.

But if comparison of genetic distances is desired and if sequences contain a certain amount of unknown nucleotides N, or if sequences of different length are being compared, it is important to use complete deletion for estimating genetic distances.

References

Aguileta G, Bielawski JP, Yang Z (2006). Evolutionary rate variation among vertebrate beta globin genes: Implications for dating gene family duplication events. Gene 380: 21–29.

Akaike H (1973). Information theory as an extension of the maximum likelihood principle. Pp 267–281 in: Petrov BN, Csaki F (Eds). Second International Symposium on Information Theory. Akademiai Kiado, Budapest.

Aliabadian M, Kaboli M, Nijman V, Vences M (2009). Molecular identification of birds: performance of distance-based DNA barcoding in three genes to delimit parapatric species. PLoS ONE 4(1): e4119. doi:10.1371/journal.pone.0004119.

Allen ES, Omland KE (2003). Novel intron phylogeny (ODC) supports plumage convergence in orioles (*Icterus*). Auk, 120: 961–969.

Alström P, Ericson PGP, Olsson U, Sundberg P (2006a). Phylogeny and classification of the avian superfamily Sylvioidea. Mol Phylogenet Evol, 38: 381–397.

Alström P, Aymí R, Clement P, Dyrcz A, Gargallo G, Hawkins AFA, Madge SC, Pearson DJ, Svensson L (2006b). Family Sylviidae (Old World Warblers). Pp 616–639 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 11. Old World Flycatchers to Old World Warblers. Lynx Edicions, Barcelona.

Alström P, Rasmussen PC, Olsson U, Sundberg P (2008). Species delimitation based on multiple criteria: the Spotted Bush Warbler *Bradypterus thoracicus* complex (Aves, Megaluridae). Zool J Linn Soc 154: 291–307.

Alström P, Fjeldså J, Fregin S, Olsson U (2011a). Gross morphology betrays phylogeny - the Scrub Warbler *Scotocerca inquieta* is not a cisticolid. Ibis 153: 87–97.

Alström P, Fregin S, Norman JA, Ericson PGP, Christidis L, Olsson U (2011b). Multilocus analysis of a taxonomically densely sampled dataset reveal extensive non-monophyly in the avian family Locustellidae. Mol Phylogenet Evol 58: 513–526.

Ames PL (1971). The morphology of the syrinx in passerine birds. Bull Peabody Mus Nat Hist 95: 151–262.

Anderson FE, Swofford DL (2004). Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. Mol Phylogenet Evol 33: 440–451.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002). Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annu Rev Ecol Syst 33: 707–740.

Backström N, Fagerberg S, Ellegren H (2008). Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. Mol Ecol 17: 964–980.

Bairlein F (2006). Family Sylviidae (Old World Warblers). Pp 492–575 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 11. Old World Flycatchers to Old World Warblers. Lynx Edicions, Barcelona.

Baker JM, Lopez-Medrano E, Navarro-Siguenza AG, Rojas-Soto OR, Omland KE (2003). Recent speciation in the orchard oriole group: divergence of *Icterus spurius spurius* and *Icterus spurius fuertesi*. Auk 120(3): 848–859.

Barker FK, Barrowclough GF, Groth JG (2002). A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. Proc R Soc B 269: 295–308.

Barker FK, Cibois A, Schikler P, Feinstein J, Cracraft J (2004). Phylogeny and diversification of the

largest avian radiation. Proc Natl Acad Sci USA 101: 11040–11045.

Bates GL (1930). Handbook of the birds of West Africa. Bale Sons and Danielson, London.

Bauer H-G, Bezzel E, Fiedler W (2005). Das Kompendium der Vögel Mitteleuropas. 2nd Edition. Aula-Verlag, Wiebelsheim.

Benzoni F, Stefani F, Pichon M, Galli P (2010). The name game: morpho-molecular species boundaries in the genus *Psammocora* (Cnidaria, Scleratinia). Zool J Linn Soc 160: 421–456.

Beresford P, Barker FK, Ryan PG, Crowe TM (2005). African endemics span the tree of songbirds (Passeri): molecular systematics of several evolutionary 'enigmas'. Proc R Soc Lond B 272: 849–858.

Bergsten J (2005). A review of long-branch attraction. Cladistics 21: 163–193.

Bertelli S, Giannini NP (2005). A phylogeny of extant penguins (Aves: Sphenisciformes) combining morphology and mitochondrial sequences. Cladistics 21: 209–239.

Bock WJ (1994). History and nomenclature of avian family-group names. Bull Am Mus Nat Hist 222: 1–283.

Bonfield JK, Smith KF, Staden R (1995). A new DNA sequence assembly program. Nucleic Acids Res 24: 4992–4999.

Bradley RD, Baker RJ (2001). A test of the genetic species concept: cytochrome-b sequences and mammals. J Mammal 82(4): 960–973.

Burbrink FT, Lawson R, Slowinski JB (2000). Mitochondrial DNA phylogeography of the polytypic North American Rat Snake (*Elaphe obsoleta*): a critique of the subspecies concept. Evolution 54(6): 2107–2118.

Cagnon C, Lauga B, Hémery G, Mouchès C (2004). Phylogeographic differentiation of storm petrels (*Hydrobates pelagicus*) based on cytochrome b mitochondrial DNA variation. Mar Biol 145: 1257–1264.

Capa M, Bybee DR, Bybee SM (2010). Establishing species and species boundaries in Sabellastarte Krøyer, 1856 (Annelidae: Sabellidae): an integrative approach. Org Divers Evol 10: 351–371.

Chesser RT (2004). Molecular systematics of New World suboscine birds. Mol Phylogenet Evol 32: 11–24.

Cibois A, Pasquet E, Schulenberg TS (1999). Molecular systematics of the Malagasy babblers (Passeriformes: Timaliidae) and warblers (Passeriformes: Sylviidae), based on cytochrome b and 16S rRNA sequences. Mol Phylogenet Evol 13: 581–595.

Cibois A, Slikas B, Schulenberg TS, Pasquet E (2001). An endemic radiation of Malagasy songbirds is revealed by mitochondrial DNA sequence data. Evolution 55: 1199–1206.

Cibois A (2003a): Mitochondrial DNA phylogeny of babblers (Timaliidae). Auk 120: 35–54.

Cibois A (2003b): Sylvia is a babbler: taxonomic implications for the families Sylviidae and Timaliidae. BOC 123: 257–261.

Cibois A, Cracraft J (2004). Assessing the passerine "Tapestry": phylogenetic relationships of the Muscicapoidea inferred from nuclear DNA sequences. Mol Phylogenet Evol 32: 264–273.

Cibois A, David N, Gregory SMS, Pasquet E (2010). Bernieridae (Aves: Passeriformes): a family-group name for the Malagasy sylvioid radiation. Zootaxa 2554: 65–68.

Cibois A, Beadell JS, Graves GR, Pasquet E, Slikas B, Sonsthagen S, Thibault J-C, Fleischer RC (2011). Charting the course of reed-warblers across the Pacific islands. J Biogeogr 38: 1963–1975.

Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN (2007). DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Mol Ecol Notes 7: 184–190.

Clements JF, Schulenberg TS, Iliff MJ, Sullivan BL, Wood CL, Roberson D (2011). The Clements checklist of birds of the world: Version 6.6. Downloaded from

http://www.birds.cornell.edu/clementschecklist/downloadable-clements-checklist

Collar NJ, Robson C (2007). Family Timaliidae (Babblers). Pp 70–133 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 12. Picathartes to Tits and Chickadees. Lynx Edicions, Barcelona.

Collura RV, Stewart C-B (1995). Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. Nature 378: 485-489.

Collins FH, Paskewitz SM (1996). A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic *Anopheles* species. Insect Mol Biol 5: 1–9.

Cracraft J, Barker FK, Braun M, Harshman J, Dyke GJ, Feinstein J, Stanley S, Cibois A, Schikler P, Beresford P, García-Moreno J, Sorenson MP, Yuri T, Mindell DP (2004). Phylogenetic relationships among modern birds (Neornithes). Toward an avian tree of life. Pp 468–489 in: Cracraft J, Donoghue MJ (Eds). Assembling the Tree of Life. Oxford University Press, Oxford.

Cramp S, Perrins CM (Eds) (1993). The Birds of the Western Palearctic. Vol VII. Oxford University Press, Oxford.

Dalén L, Götherström A, Angerbjörn A (2004). Identifying species from pieces of faeces. Conserv Genet 5(1): 109–111.

Davison A, Blackie RLE, Scothern GP (2009). DNA barcoding of stylommatophoran land snails: a test of existing sequences. Mol Ecol Resour 9: 1092–1101.

de Juana E, Suárez F, Ryuan PG (2004). Family Alaudidae (Larks). Pp 496–541 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 9. Cotingas to Pipits and Wagtails. Lynx Edicions, Barcelona.

del Hoyo J, Elliott A, Christie DA eds (2005). Handbook of the Birds of the World. Vol 10. Cuckoo-shrikes to Thrushes. Lynx Edicions, Barcelona.

del Hoyo J, Elliott A, Christie DA eds (2006). Handbook of the Birds of the World. Vol 11. Old world flycatchers to Old World Warblers. Lynx Edicions, Barcelona.

del Hoyo J, Elliott A, Christie DA eds (2007). Handbook of the Birds of the World. Vol 12. Picathartes to Tits and Chickadees. Lynx Edicions, Barcelona.

Dickinson EC (Ed) (2003). The Howard and Moore Complete Checklist of the Birds of the World, 3rd Edition Princeton University Press, Princeton, New Jersey.

Drovetski SV, Zink RM, Fadeev IV, Nesterov EV, Koblik EA, Red'kin YA, Rohwer S (2004). Mitochondrial phylogeny of *Locustella* and related genera. J Avian Biol 35: 105–110.

Drummond AJ, Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC 7: 214. Available from: <http://beast.bio.ed.ac.uk/Tracer>.

Dyke G, Kaiser G (Eds) (2011). Living dinosaurs: the evolutionary history of modern birds. John Wiley & Sons, Ltd, Chichester, UK.

Edwards SV, Wilson AC (1990). Phylogenetically informative length polymorphism and sequence variability in mitochondrial DNA of Australian songbirds (*Pomatostomus*). Genetics 126: 695–711.

Ekrem T, Stur E, Hebert PD (2010). Females do count: documenting Chironomidae (Diptera) species diversity using DNA barcoding. Org Divers Evol 10: 397–408.

Elias M, Hill RI, Willmott KR, Dasmahapatra KK, Brower AVZ, Mallet J, Jiggins CD (2007). Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc R Soc B 274: 2881–2889.

Ericson PGP, Johansson US, Parsons TJ (2000). Major divisions of oscines revealed by insertions in the nuclear gene c-myc: a novel gene in avian phylogenetics. Auk 117: 1077–1086.

Ericson PGP, Christidis L, Cooper A, Irestedt M, Jackson J, Johansson US, Norman JA (2002). A

Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. Proc Roy Soc Lond B 269: 235–241.

Ericson PGP, Johansson US (2003). Phylogeny of Passerida (Aves: Passeriformes) based on nuclear and mitochondrial sequence data. Mol Phylogenet Evol 29: 126–138.

Erixon P, Svennblad B, Britton T, Oxelman B (2003). Reliability of Bayesian Posterior Probabilities and Bootstrap Frequencies in Phylogenetics. Syst Biol 52(5): 665–673.

Feduccia A (1974). Morphology of the bony stapes in New and Old World suboscines: new evidence for common ancestry. Auk 91: 427–429.

Feduccia A (1996). The origin and evolution of birds. Yale University Press, .New Haven and London.

Felsenstein J (1978). Cases in which parsimony and compatibility methods will be positively misleading. Syst Zool 27: 401–410.

Felsenstein J (1981). Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. J Mol Evol 17: 368–376.

Fernando P, Pfrender ME, Encalada SE, Lande R (2000). Mitochondrial DNA variation, phylogeography and population structure of the Asian elephant. Heredity 4: 362–32.

Fishpool LDC, Tobias JA (2005). Family Pycnonotidae (Bulbuls). Pp 124–167 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 10. Cuckoo-shrikes to Thrushes. Lynx Edicions, Barcelona.

Fjeldså J, Zuccon D, Irestedt M, Johansson US, Ericson PGP (2003). *Sapayoa aenigma*: a New World representative of 'Old World suboscines'. Proc R Soc Lond B 270: 238–241 (Suppl).

Forbes WA (1882). Contributions to the anatomy of passerine birds. Part 6. On *Xenicus* and *Achantisitta* as types of a new family (Xenicidae) of mesomyodian Passeres from New Zealand. Proc Zool Soc Lond 1882: 569–571.

Freeman S, Herron JC (2001). Evolutionary Analysis. 2nd Edition. Prentice Hall, Inc., Upper Saddle River, New Jersey.

Friesen VL, Congdon BC, Kidd MG, Birt TP (1999). Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. Mol Ecol 8: 2141–2152.

Fuchs J, Fjeldså J, Bowie RCK, Voelker G, Pasquet E (2006). The African warbler genus *Hyliota* as a lost lineage in the oscine songbird tree: molecular support for an African origin of the Passerida. Mol Phylogenet Evol 39: 186–197.

Fürbringer M (1888). Untersuchungen zur Morphologie und Systematik der Vögel, zugleich ein Beitrag zur Anatomie der Stütz- und Bewegungsorgane, 2 Volumes. Amsterdam, Holland: Van Holkema TJ.

Garnery L, Cornuet J-M, Solignac M (1992). Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. Mol Ecol 1(3): 145–154.

Gelang M, Cibois A, Pasquet E, Olsson U, Alström P, Ericson PGP (2009). Phylogeny of babblers (Aves, Passeriformes): major lineages family limits and classifications. Zool Scr 38: 225–236.

Gellissen GJ, Bradfield Y, White BN, Wyatt GR (1983). Mitochondrial DNA sequences in the nuclear genome of a locust. Nature 301: 631-634.

Gill BJ (2004). Family Acanthisittidae (New Zealand Wrens). Pp 464–471 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 9. Cotingas to Pipits and Wagtails. Lynx Edicions, Barcelona.

Gill F, Donsker D (Eds) (2011). IOC World Bird Names (v 2.20). Available at: http://www.worldbirdnames.org/ (Accessed November 2011).

Gill FB, Sheldon FH (1991). The Birds Reclassified. Science, 17 May 1991: 1003–1005.

Gomez A, Wright PJ, Lunt DH, Cancino JM, Carvalho GR, Hughes RN (2007). Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan taxon. Proc R Soc B 274: 99–207.

Greenstone MH, Rowley DL, Heimbach U, Lundgren JG, Pfannenstiel RS, Rehner SA (2005). Barcoding generalist predators by polymerase chain reaction: carabids and spiders. Mol Ecol 14: 3247–3266.

Griffiths CS (1997). Correlation of Functional Domains and Rates of Nucleotide Substitution in Cytochrome b. Mol Phylogenet Evol 7(3): 352–365.

Griffiths CS, Barrowclough GF, Groth JG, Mertz L (2004). Phylogeny of the Falconidae (Aves): a comparison of the efficacy of morphological, mitochondrial, and nuclear data. Mol Phylogenet Evol 32: 101–109.

Groth JG (1998) Molecular phylogenetics of finches and sparrows: consequences of character state removal in cytochrome b sequences. Mol Phylogenet Evol 10: 377–390.

Groth JC, Barrowclough GF (1999). Basal divergencies in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol Phylogenet Evol 12: 115–123.

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

Haase M, Marshall B, Hogg I (2007). Disentangling causes of disjunction on the South Island of New Zealand: the Alpine fault hypothesis of vicariance revisited. Biol J Linn Soc 91: 361–374.

Hackett SJ (1996). Molecular Phylogenetics and Biogeography of Tanagers in the Genus *Ramphocelus* (Aves). Mol Phylogenet Evol 5(2): 368–382.

Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T (2008). A phylogenomic study of birds reveals their evolutionary history. Science 320: 1763–1768.

Haffer J (1991). Familie Sylviidae-Zweigsänger (Grasmücken und Verwandte). Pp 11–21 and 208–216 in: von Blotzheim GUN, Bauer KM (Eds). Handbuch der Vögel Mitteleuropas. Vol 12. Aula-Verlag, Wiesbaden.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95–98.

Harrap S (2008). Family Aegithalidae (Long-tailed tits). Pp 76–93 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 13. Penduline-tits to Shrikes. Lynx Edicions, Barcelona.

Harshman J (2006): The Tree of Life Web Project http://tolweborg/Sylvioidea/67276/20060802 (Sylvioidea Version 02 August 2006; under construction).

Hasegawa M, Kishino H, Yano T (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160-174.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a). Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270: 313–321.

Hebert PDN, Ratnasingham S, DeWaard JR (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B Biol Sci 270: S596–S599.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004). Identification of birds through DNA barcodes. PloS Biol 2(10): e312.

Helbig AJ, Seibold I, Martens J, Wink M (1995). Genetic differentiation and phylogenetic relationships of Bonelli´s Warbler *Phylloscopus bonelli* and Green Warbler *P. nitidus*. J Avian Biol 26: 139–153.

Helbig AJ, Martens J, Seibold I, Henning F, Schottler B, Wink M (1996). Phylogeny and species limits

in the palaearctic chiffchaff *Phylloscopus collybita* complex: mitochondrial genetic differentiation and bioacoustic evidence. Ibis 138: 650–666.

Helbig AJ, Seibold I (1999). Molecular Phylogeny of Palearctic–African *Acrocephalus* and *Hippolais* Warblers (Aves: Sylviidae). Mol Phylogenet Evol 11: 246–260.

Helbig AJ, Kocum A, Seibold I, Braun MJ (2005). A multi-gene phylogeny of aquiline eagles (Aves: Accipitriformes) reveals extensive paraphyly at the genus level. Mol Phylogenet Evol 35: 147–164.

Hendy MD, Penny D (1989). A framework for the quantitative study of evolutionary trees. Syst Zool 38: 297–309.

Heslewood MM, Elphinstne MS, Tidemann SC, Baverstock PR (1998). Myoglobin intron variation in the Gouldian Finch *Erythura gouldiae* assessed by temperature gradient gel electrophoresis. Electrophoresis 19: 142–151.

Hickerson MJ, Gilchrist MA, Takebayashi N (2003). Calibrating a molecular clock from phylogeographic data: moments and likelihood estimators. Evolution 57: 2216–2225.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001). Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. Science 294: 2310–2314.

Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754–755.

Hugenholtz P, Pace NR (1996). Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. Trends Biotechnol 14: 190–197.

Hung G-C, Chilton NB, Beveridge I, Zhu XQ, Lichtenfels JR, Gasser RG (1999). Molecular evidence for cryptic species within *Cylicostephanus minutus* (Nematoda: Strongylidae). Int J Parasitol 29: 285–291.

Hurvich CM, Tsai C-L (1989). Regression and time series model selection in small samples. Biometrika 76: 297–307.

Irestedt M, Johansson US, Parsons TJ, Ericson PGP (2001). Phylogeny of major lineages of suboscines (Passeriformes) analysed by nuclear DNA sequence data. J Avian Biol 32: 15–25.

Irestedt M, Ohlson JI, Zuccon D, Källersjö M, Ericson PGP (2006). Nuclear DNA from old collections of avian study skins reveals the evolutionary history of the Old World suboscines (Aves, Passeriformes). Zool Scr 35: 567–580.

Jiang F, Miao Y, Liang W, Ye H, Liu H, Liu B (2009). The complete mitochondrial genomes of the whistling duck (*Dendrocygna javanica*) and black swan (*Cygnus atratus*): dating evolutionary divergence in Galloanserae. Mol Biol Rep 37(6): 3001–30015.

Jobb G, von Haeseler A, Strimmer K (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4: 1471–2148.

Jobb G (2008). TREEFINDER version of October 2008. Munich, Germany. Distributed by the author at <www.treefinder.de>.

Johansson U, Fjeldså J, Bowie R (2008). Phylogenetic relationships within Passerida (Aves: Passeriformes): a review and a new molecular phylogeny based on three nuclear intron markers. Mol Phylogenet Evol 48: 858–876.

Johnson JB, Omland KS (2003). Model selection in ecology and evolution. Trends Ecol Evol 19: 101–108.

Johnson NK, Cicero C (2004). New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. Evolution 58: 1122–1130.

Jukes TH, Cantor CR (1969). Evolution of protein molecules. Pp 21–132 in: Munro H (Ed). Mammalian Protein Metabolism. Academic Press.

Katho K, Misawa K, Kuma K-I, Miyata T (2002). MAFFT: a novel method for rapid multiple sequence

alignment based on fast Fourier transform. Nucl Acids Res 30: 3059–3066.

Kemble RJ, Mans RM, Gabay-Laughnan S, Laughnan JR (1983). Sequences homologous to episomal mitochondrial DNAs in the maize nuclear genome. Nature 304: 744-747.

Kennerley PR, Leader PJ (1992). The identification, status and distribution of small *Acrocephalus* warblers in Eastern China. Hong Kong Bird Report 1991: 143–187.

Klicka J, Zink RM, Barlow JC, McGillivray WB, Doyle TJ (1999). Evidence supporting the recent origin and species status of the Timberline Sparrow. Condor 101: 577–588.

Kluge AG, Farris JS (1969). Quantitative phyletics and the evolution of Anurans. Syst Zool 18: 1–32.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci USA 86: 6196–6200.

Kullen MJ, Sanozky-Dawes RB, Crowell DC, Klaenhammer TR (2000). Use of the DNA sequence of variable regions of the 16S rRNA gene for rapid and accurate identification of bacteria in the *Lactobacillus acidophilus* complex. J Appl Microbiol 89: 511–516.

Lanave C, Preparata C, Saccone C, Serio G (1984). A new method for calculating evolutionary substitution rates. J Mol Evol 20: 86–93.

Langhoff P, Authier A, Buckley TR, Dugdale JS, Rodrigos A, Newcomb RD (2009). DNA barcoding of the endemic New Zealand leafroller moth genera, *Ctenopseustis* and *Planotortrix*. Mol Ecol Resour 9: 691–698.

Leisler B (1980). Morphological aspects of ecological specialization in bird genera. Oekol Vögel 2: 199–220.

Leisler B, Heidrich P, Schulze-Hagen, K, Wink M (1997). Taxonomy and phylogeny of reed warblers (genus *Acrocephalus*) based on mtDNA sequences and morphology. J Ornithol 138: 469–496.

Leisler B, Ley HW, Winkler H (1989). Habitat, behaviour and morphology of *Acrocephalus* warblers: an integrated analysis. Ornis Scand 20: 181–186.

Lemmon AR, Moriarty EC (2004). The importance of proper model assumption in Bayesian phylogenetics. Syst Biol 53: 265–277.

Lin C-P, Danforth BN (2004). How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol Phylogenet Evol 30: 686–702.

Livezey BC, Zusi RL (2007) Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and Discussion. Zool J Linn Soc 149: 1–95.

Loader SP, Gower DJ, Ngalason W, Menegon M (2010). Three new species of *Callulina* (Amphibia: Anura: Brevicipitidae) highlight local endemism and conservation plight of Africa′ s Eastern Arc forests. Zool J Linn Soc 160: 496–514.

Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (1994). Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol 39: 174–190.

Mackie IM, Pryde SE, Gonzales-Sotelo C, Medina I, Perez-Martino R, Quinterio J (1999). Challenges in the identification of species of canned fish. Trends Food Sci Tech 10: 9–14.

Markolf M, BrameierM , Kappeler PM (2011). On species delimitation: Yet another lemur species or just genetic variation? BMC Evolutionary Biology 2011, 11:216 doi:10.1186/1471-2148-11-216.

Marova IM, Valchuk OP, Kvartalynov PV, Ivanitskij VV (2005). On the taxonomic position and evolutionary interrelations of the Thick-billed Warbler, *Phragmaticola aedon* (based on ecological an ethological data). Alauda 73: 308.

Meier R, Shiyang K, Vaidya G, Peter KLNg (2006). DNA barcoding and taxonomy in Diptera: a tale of

high intraspecific variability and low identification success. Syst Biol 55(5): 715–728.

Meier R, Zhang G, Ali F (2008). The use of mean instead of smallest interspecific distances exaggerates the size of the ''barcoding gap'' and leads to misidentification. Syst Biol 57(5): 809–813.

Meise W (1976). Pp 207–216 in: Proceedings of the 16th International Ornithological Congress. Vol 1974.

Mayr G (2011). Metaves, Mirandornithes, Strisores, and other novelties – a critical review of the higher-level phylogeny of neornithine birds. J Zool Syst Evol Res 49: 58–76.

Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16: 1215.

Mindell DP (1992). DNA-DNA Hybridisation and avian phylogeny Syst Biol 41(1): 126–134.

Morrison DA, Ellis JT (1997). Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of apicomplexa. Mol Biol Evol 14: 428–441.

Nanney DL (1982). Genes and phenes in Tetrahymena. Bioscience 32: 783–788.

Newman CE, Feinberg JA, Rissler LJ, Burger J, Shaffer HB (2012). A new species of leopard frog (Anura: Ranidae) from the urban northeastern US. Mol Phylogenet Evol 63(2): 445–455.

Nguembock B, Fjeldså J, Tillier A, Pasquet E (2007). A phylogeny for the Cisticolidae (Aves: Passeriformes) based on nuclear and mitochondrial DNA sequence data, and a re-interpretation of an unique nest-building specialization. Mol Phylogenet Evol 42: 272–286.

Nylander JAA (2004): MrModeltest v2 Program distributed by the author Evolutionary Biology Centre Uppsala University. [http://wwwabcse/~nylander/]

Notredame C, Higgins D, Heringa J (2000). T–Coffee: a novel method for multiple sequence alignments. J Mol Biol 302: 205–217. Available from: <http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi>.

Olsson U, Alström P, Ericson PGP, Sundberg P (2005). Non-monophyletic taxa and cryptic species – evidence from a molecular phylogeny of leaf-warblers (*Phylloscopus*, Aves). Mol Phylogenet Evol 36: 261–276.

Pace NR, Stahl DA, Lane DJ, Olsen GJ (1986). The analysis of natural microbial populations by ribosomal RNA sequences. Adv Microb Eecol 9: 1–55.

Palma RE, Rodriguez-Serrano E, Rivera-Milla E, Hernandez CE, Salazar-Bravo J, Carma MI, Belmar-Lucero S, Gutierrez-Tapia P, Zeballos H, Yates TL (2010). Phylogenetic relationships of the pygmy rice rats of the genus *Oligoryzomys* Bangs, 1900 (Rodentia: Sigmodontinae). Zool J Linn Soc Lond 160: 551–566.

Parkin DT, Collinson M, Helbig AJ, Knox AG, Sangster G, Svensson L (2004). Species limits in *Acrocephalus* and *Hippolais* warblers from the Western Palearctic. Brit Birds 97: 276–299.

Posada D, Crandall KA (2001). Selecting the best-fit model of nucleotide substitution. Syst Biol 50(4): 580–601.

Posada D, Buckley TR (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 53: 793–808.

Posada D (2008). jModelTest: phylogenetic model averaging. Mol Biol Evol 25(564): 1253–1256.

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995). Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines Proc Natl Acad Sci USA 92: 7759–7763.

Quinn TW, White BN (1987). Analysis of DNA sequence variation. In: Cooke F, Buckley PA (Eds). Avian genetics. Academic Press, London.

Ratnasingham S, Hebert PDN (2007). BOLD: the barcoding of life data system (www.barcodinglife.org). Mol Ecol Notes 7: 355–364, doi:10.1111/j.1471-8286.2006.01678.

Robson C (2007). Family Paradoxornithidae (Parrotbills). Pp 292–307 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 12. Picathartes to Tits and Chickadees. Lynx Edicions, Barcelona.

Rodríguez J, Oliver L, Marín A, Medina R (1990). The general stochastic model of nucleotide substitution. J Theo Biol 142: 485–501.

Roe AD, Sperling FAH (2007). Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. Mol Phylogenet Evol 44: 325–345.

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

Ryan PG (2006). Family Cisticolidae (Cisticolas and allies). Pp 378–419 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 11. Old World Flycatchers to Old World Warbles. Lynx Edicions, Barcelona.

Sangster G, Collinson JM, Knox AG, Parkin DT, Svensson L (2010). Taxonomic recommendations for British birds: Sixth report. Ibis 152: 180–186.

Sarich VM, Schmid CW, Marks J (1989). DNA hybridization as a guide to phylogenies: a critical analysis. Cladistics 5: 3–32.

Saunders GW (2005). Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. Phil Trans R Soc B 360(1462): 1879-1888.

Schenk JJ, Hufford L (2010). Effects of substitution models on divergence time estimates: simulations and an empirical study of model uncertainty using cornales. Syst Bot 35(3): 578–592.

Schönwetter M (1979). Handbuch der Oologie. Band II. Akademie-Verlag, Berlin.

Sefc KM, Payne RB, Sorenson MD (2003). Phylogenetic relationships of African sunbird-like warblers: Moho (*Hypergerus atriceps*), Green Hylia (*Hylia prasina*) and Tit-hylia (*Pholidornis rushiae*). Ostrich 74: 8–17.

Sheldon FH, Gill FB (1996). A reconsideration of songbird phylogeny, with emphasis on the evolution of titmice and their sylvioid relatives. Syst Biol 45: 473–495.

Sibley CG, Ahlquist JE (1982). The relationships of the wrentits as indicated by DNA-DNA Hybridization. Condor 84: 40–44.

Sibley CG, Ahlquist JE (1990). Phylogeny and Classification of Birds – A Study in Molecular Evolution. Yale University Press, New Haven and London.

Sibley CG, Monroe BL Jr (1990): Distribution and Taxonomy of Birds of the World. Yale University Press, New Haven and London.

Slade RW, Moritz C, Heideman A, Hale PT (1993). Rapid assessment of single-copy nuclear DNA variation in diverse species. Mol Ecol 2: 359–373.

Song H, Buhay JE, Whiting MF, Crandall KA (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. P Natl Acad Sci USA 105: 13486–13491.

Sorenson MD, Quinn TW (1998). Numts: a challenge for avian systematics and population biology. Auk 115: 214–221.

Spellman GM, Cibois A, Moyle RG, Winker K, Barker FK (2008). Clarifying the systematic of an enigmatic avian lineage: what is a bombycillid? Mol Phylogenet Evol 49: 1036–1040.

Springer M, Krajewksi C (1989). DNA hybridization in animal taxonomy: a critique from first principles. Q Rev Biol 64: 291–318.

Steane DA, Conod SN, Jones RC, Vaillancourt RE, Potts BM (2006). A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite markers and quantitative traits. Tree Genet Genomes 2(1): 30–38.

Sugiura N (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. Commun Stat Theory Methods A7: 13–26.

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

Svensson L, Mullarney K, Zetterström D (2009). Birds of Europe. 2nd Edition. Princeton University Press, Princeton, New Jersey.

Swofford DL, Olsen GP, Waddell PJ, Hillis DM (1996). Phylogenetic inference. Pp 407–514 in: Hillis DM, Moritz C, Mable BK (Eds). Molecular Systematics. 2nd Edition. Sinauer, Sunderland, MA.

Swofford DL (2003). PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4: Sinauer Associates Sunderland Massachusetts.

Tamura K, Nei M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512–526.

Tavaré S (1986). Some probabilistic and statistical problems on the analysis of DNA sequences. Lect Mater Life Sci 17: 57–86.

Treplin S, Siegert R, Bleidorn C, Thompson HS, Fotso R, Tiedemann R (2008). Molecular phylogeny of songbirds (Aves: Passeriformes) and the relative utility of common nuclear marker loci. Cladistics 24(3): 328–349.

Turner AK (2004). Family Hirundinidae (Larks). Pp 602–640 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 9. Cotingas to Pipits and Wagtails. Lynx Edicions, Barcelona.

van Balen S (2008). Family Zosteropidae (White-eyes). Pp 402–443 in: del Hoyo J, Elliott A, Christie DA (Eds). Handbook of the Birds of the World. Vol 13. Penduline-tits to Shrikes. Lynx Edicions, Barcelona.

Voous KH (1975). An aberrant Reed Warbler, or: on the inequality of genera in birds. Ardeola 21(2): 977–985.

Watson GE, Traylor MA, Mayr E (1986). Family Sylviidae, Old World Warblers. Pp 3–294 in: Mayr E, Cottrell GW (Eds). Check-list of the Birds of the World. Vol 4. Mus of Comp Zool, Cambridge, MA.

Wesson DM, McLain DK, Oliver JH, Piesman J, Collins F (1993). Investigation of the validity of species status of *Ixodes dammini* (Acari: Ixodidae) using rDNA. Proc Natl Acad Sci USA 90: 10221–10225.

Wiemers M, Fiedler K (2007). Does the DNA barcoding gap exist – a case study in blue butterflies. Front Zool 4(8) doi:10.1186/1742-9994-4-8.

Wolfe KH, Sharp PM, Li WH (1989). Mutation rates differ among regions of the mammalian genome. Nature 337: 283-285.

Wolters, HE (1975-82). Die Vogelarten der Erde. Paul Parey, Hamburg and Berlin.

Wright RM, Cummings DJ (1983). Integration of mitochondrial gene sequences within the nuclear genome during senescence in a fungus. Nature 302: 86–88.

Yang Z, Goldman N, Friday A (1994). Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Mol Biol Evol 11: 316–324.

Yang Z (1995). A Space-Time Process Model for the Evolution of DNA Sequences. Genetics 139: 993–1005.

Zhang D-X, Hewitt GM (1996). Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol Evol 11: 247–251.

Zhang S, Yang L, Yang X, Yang J (2007). Molecular phylogeny of the yuhinas (Sylviidae: *Yuhina*): a paraphyletic group of babblers including *Zosterops* and Philippine *Stachyris*. J Ornithol 148: 417–426.

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

Own contribution to manuscripts

I Multi-locus phylogeny of the family Acrocephalidae (Aves: Passeriformes) – The traditional taxonomy overthrown

Data acquisition: 75%; Analyses, Figures/Tables: 90%, Text: 70%

II Multilocus analysis of a taxonomically densely sampled dataset reveal extensive non-monophyly in the avian family Locustellidae

Data acquisition, Analyses, Figures/Tables, Text: 15%

III Pitfalls in comparisons of genetic distances: A case study of the avian family Acrocephalidae

Data acquisition, Analyses: 80%, Figures/Tables: 90%, Text: 80%

IV New insights into family relationships within the avian superfamily Sylvioidea (Passeriformes) based on seven molecular markers

Data acquisition: 75%; Analyses, Figures/Tables: 90%, Text: 80%

Molecular Phylogenetics and Evolution 52 (2009) 866–878

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Multi-locus phylogeny of the family Acrocephalidae (Aves: Passeriformes) – The traditional taxonomy overthrown

Silke Fregin [a,*], Martin Haase [a], Urban Olsson [b], Per Alström [c,d]

[a] Vogelwarte Hiddensee, Zoological Institute and Museum, Ernst Moritz Arndt University of Greifswald, 17489 Greifswald, Germany
[b] Department of Zoology, University of Göteborg, Box 463, SE-405 30 Göteborg, Sweden
[c] Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, SE-750 07 Uppsala, Sweden
[d] Department of Vertebrate Zoology and Molecular Systematics Laboratory, Swedish Museum of Natural History, P.O. Box 50007, SE-104 05 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 9 January 2009
Revised 8 April 2009
Accepted 15 April 2009
Available online 23 April 2009

Keywords:
Aves, Passeriformes
Sylvioidea
Acrocephalus
Chloropeta
Hippolais
Nesillas
Cytochrome *b*
Lactate dehydrogenase intron 3
Ornithine decarboxylase introns 6 and 7
Myoglobin intron 2

ABSTRACT

We present the first study of the warbler family Acrocephalidae based on one mitochondrial and three nuclear DNA loci, in total ~2900 bp, including most or all of the species in three (*Acrocephalus*, *Hippolais* and *Chloropeta*) of the four genera and one species in the fourth genus (*Nesillas*) in this family. All three genera were suggested to be non-monophyletic, although the non-monophyly of *Acrocephalus* is not fully convincingly demonstrated. Six major clades were found, which agreed largely with the results from two earlier mitochondrial studies, and for which the names *Hippolais*, *Iduna*, *Acrocephalus*, *Calamocichla*, *Notiocichla* and *Calamodus* have been used. However, the results also revealed some new constellations, due to better resolution of deeper nodes and the inclusion of more taxa. The taxonomic implications are discussed.

1. Introduction

The warbler family Acrocephalidae consists of the four genera *Acrocephalus* (“reed warblers”), *Hippolais* (“tree warblers”), *Chloropeta* (“yellow warblers”) and *Nesillas* (“brush warblers”) (Johansson et al., 2008). In total, 53 species and 82 subspecies are currently recognized, with *Acrocephalus* being the largest (37 spp./67 sspp.) and *Chloropeta* the smallest (3 spp./6 sspp.) genus (del Hoyo et al., 2006). The members of this family breed widely across the Old World and Australasia: *Acrocephalus* in Eurasia (17 spp.; mainly Palearctic), Africa/Madagascar/Seychelles/Mascarene islands (7 spp.) and Australia/Polynesia (13 spp.); *Hippolais* exclusively in the Palearctic; *Chloropeta* in Subsaharan Africa; and *Nesillas* in Madagascar and on the Comoro islands (del Hoyo et al., 2006). All Eurasian species are wholly or partly migratory, with a rough trend that species breeding in the Western Palearctic winter in Africa, while those breeding further east winter in Southern Asia; the other species are mainly sedentary (Cramp, 1992; del Hoyo et al., 2006). Most species of Acrocephalidae occur in wet habitats, such as reedbeds and other marshland vegetation, but some, e.g. all *Hippolais* and *Nesillas* species, live in drier habitats (del Hoyo et al., 2006). Nearly all species have rather plain, nondescript plumage, with brownish or greyish uppersides in various shades, and whitish underparts with more or less distinct brownish or buffish hues. A few *Acrocephalus*, two *Hippolais* and all *Chloropeta* species are greenish-tinged above and have pale to bright yellow underparts. Four *Acrocephalus* species stand out by showing dark streaking. There is considerable variation in size between different species, especially in *Acrocephalus* and *Hippolais*. The songs are highly varied and in some species contain a high degree of mimicry, and generally differ more than morphology between closely related species (Cramp, 1992; del Hoyo et al., 2006).

The relationships among the genera in Acrocephalidae sensu Johansson et al. (2008) and their relationships to other taxa have long been debated based on morphology and oology (Cramp, 1992; Dickinson, 2003; Haffer, 1991; Hartert, 1909; Meise, 1976; Schönwetter, 1979; Voous, 1975; Watson et al., 1986; Wolters, 1982) as well as DNA (e.g. Alström et al., 2006; Beresford et al., 2005; Helbig and Seibold, 1999; Johansson et al., 2008; Leisler

* Corresponding author. Fax: +49 3834 864332.
E-mail addresses: silke.fregin@uni-greifswald.de (S. Fregin), martin.haase@uni-greifswald.de (M. Haase), urban.olsson@zool.gu.se (U. Olsson), Per.Alstrom@artdata.slu.se (P. Alström).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.04.006

et al., 1997; Sibley and Ahlquist, 1990). The uncertainty is due to their morphological similarities (e.g. Cramp, 1992; del Hoyo et al., 2006; Kennerley and Leader, 1992), convergent adaptations to similar habitats, and rapid radiation (e.g. del Hoyo et al., 2006; Haffer, 1991; Leisler, 1980; Leisler et al., 1989; Parkin et al., 2004; Sibley and Ahlquist, 1990). On morphological grounds, the "reed warbler group" *sensu lato* (s.l.) has been characterized in various ways. *Chloropeta* was suggested to be a member by several authors (Erard et al., 1997; Watson et al., 1986; Wolters, 1982; Table 1), while Haffer (1991) proposed *Locustella* to be included. In the DNA-DNA hybridization study by Sibley and Ahlquist (1990), *Acrocephalus*, with *Hippolais* as its sister genus, was placed together with e.g. *Cettia*, *Bradypterus*, *Locustella*, *Chloropeta*, *Sylvietta* and *Phylloscopus* in the subfamily Acrocephalinae of the family Sylviidae; *Sylvietta* and *Phylloscopus* were suggested to be the nearest relatives of *Acrocephalus*/*Hippolais*. Phylogenetic analyses based on mitochondrial cytochrome *b* sequences (Helbig and Seibold, 1999; Leisler et al., 1997) rejected the hypothesis that *Locustella* belongs to the "reed warbler group" s.l. Based on sequence data from the nuclear RAG-1 and -2 exons, Beresford et al. (2005) found *Acrocephalus* in a clade with Malagasy "warblers" (*Thamnornis* and *Xanthomixis*), *Cincloramphus*, *Megalurus*, *Schoenicola* and *Bradypterus*, although the support for the inclusion of *Acrocephalus* was weak (*Hippolais* and *Chloropeta* were not included). Using one mitochondrial and one nuclear locus, Alström et al. (2006) recovered a strongly supported clade comprising *Acrocephalus*, *Hippolais* and *Chloropeta*, which they referred to as Acrocephalidae. This was sister to a clade with *Bradypterus* and *Megalurus* (denoted Megaluridae) and *Donacobius*, although the support for that sister relationship was not unambiguously strong (none of the other taxa in the Beresford et al. (2005) paper were included). The most recently published and most densely sampled study (Johansson et al., 2008) found a strongly supported clade containing *Acrocephalus*, *Hippolais* and *Nesillas* (Acrocephalidae; *Chloropeta* not included) in a well supported sister relation to a clade including e.g. *Bradypterus*, *Locustella*, *Megalurus* (Megaluridae), three genera of Malagasy "warblers" (Bernieridae), and *Donacobius*; *Nesillas* was sister to *Acrocephalus* and *Hippolais*.

Not only higher-level relationships have proved difficult to establish, but also relations within the group. The history of the classification reflects the morphological similarities and the problems of distinguishing homologous and homoplasious morphological characters, as well as overemphasis of autapomorphic characters. In the second half of the 20th century the three genera *Acrocephalus*, *Hippolais* and *Chloropeta* were split into several more genera and/or subgenera (Dickinson, 2003; Grant and Mackworth-Praed, 1940; Haffer, 1991; Helbig and Seibold, 1999; Leisler et al., 1997; Sibley and Monroe, 1990; Watson et al., 1986; Wolters, 1982; Table 1). However, there has been little consensus of how to classify these taxa, especially *Acrocephalus*. Leisler et al. (1997) and Helbig and Seibold (1999) came to nearly identical conclusions based on analyses of cytochrome *b* sequences. Neither of them supported the monophyly of the genera *Acrocephalus* and *Hippolais*. Leisler et al. (1997) discussed the possibility of splitting these genera into a number of genera, and Helbig and Seibold (1999) explicitly proposed a revised classification into a number of subgenera. Both agreed on the classification of the small striped reed warblers (including the unstriped, but otherwise morphologically and vocally similar *A. bistrigiceps*) as *Calamodus*; the other small, unstreaked, species as *Notiocichla*; the "*arundinaceus*/*stentoreus* group"

Table 1
Overview of some different taxonomic treatments of the Acrocephalidae (except *Nesillas*). Grey background, genus status; white background, subgenus status; / species not considered in study. The divisions in the leftmost column indicate the main clades found in the present study.

Taxon	Dickinson 2003	Helbig & Seibold 1999		Leisler et al. 1997		Haffer 1991		Sibley & Monroe 1990	Watson et al. 1986		Wolters 1982	
A. aedon	*Phragamaticola*	*Phragamaticola*	*Acrocephalus*	/		*Phragamaticola*	*Acrocephalus*		*Phragamaticola*	*Acrocephalus*	*Phragamaticola*	
A. melanopogon	*Acrocephalus*	*Calamodus*		*Calamodus*		*Lusciniola*		*Acrocephalus*	*Lusciniola*		*Lusciniola*	*Calamodus*
A. schoenobaenus						*Calamodus*			*Calamodus*		*Calamodus*	
A. paludicola											*Calamodyta*	
A. bistrigiceps											*Anteliocichla*	
A. tangorum		*Notiocichla*		*Notiocichla*		/			*Acrocephalus*		*Notiocichla*	*Acrocephalus*
A. concinens						*Acrocephalus*						
A. agricola												
A. scirpaceus												
A. baeticatus												
A. palustris												
A. dumetorum												
A. orinus		/		/		/		/				
A. griseldis		*Indeterminate*		*Acrocephalus*	*Acrocephalus*	*Acrocephalus*		*Acrocephalus*			*Acrocephalus*	
A. arundinaceus		*Acrocephalus*										
A. orientalis						/						
A. stentoreus				/								
A. australis				*Acrocephalus*		*Acrocephalus*		/			*Conopoderas*	
A. mendanae		/		/		/		*Acrocephalus*				
A. aequinoctialis								/				
A. taiti						*Acrocephalus*		*Acrocephalus*				
A. caffer								/				
A. atyphus								*Acrocephalus*				
A. percernis								/				
A. kerearako								*Acrocephalus*				
A. rimitarae								/				
A. familiaris						*Acrocephalus*						
A. rufescens		*Calamocichla*		*Calamocichla*		*Calamocichla*			*Calamocichla*		*Calamocichla*	
A. brevipennis												
A. newtoni												
A. gracilirostris												
A. sechellensis								*Bebrornis*	*Bebrornis*			
H. icterina	*Hippolais*	*Hippolais*	*Hippolais*	*Hippolais*		*Hippolais*		*Hippolais*	*Hippolais*		*Hippolais*	*Hippolais*
H. polyglotta												
H. languida												
H. olivetorum												
H. pallida	*Iduna*	*Iduna*		*Iduna*							*Iduna*	
H. opaca												
H. caligata												
H. rama												
C. gracilirostris	*Chloropeta*	*Chloropeta*		/		/		*Chloropeta*	*Chloropeta*		*Calamonastides*	*Chlp.*
C. natalensis		/									*Chloropeta*	
C. similis												

as *Acrocephalus sensu stricto* (s.s.); the African/Malagasy reed warblers as *Calamocichla*; the small *Hippolais* as *Iduna*; and the larger *Hippolais* as *Hippolais* (s.s.) (Table 1). The monotypic genus *Luscinia*, comprising *L. melanopogon* (Vaurie, 1959), was rejected, as this species was shown to be nested within the *Calamodus* clade. *A. aedon* was placed in a monotypic subgenus *Phragamaticola* by Helbig and Seibold (1999) (spelled as *Phragmaticola*, but see Dickinson, 2007; not included by Leisler et al., 1997). *A. griseldis* was regarded by Helbig and Seibold (1999) as a separate "indeterminate" subgenus, because of its uncertain phylogenetic position. Helbig and Seibold (1999) included only one species of *Chloropeta* (*gracilirostris*), which was shown to belong in the clade with *Acrocephalus* and *Hippolais*, albeit in an unresolved position. This species had previously been suggested to be more closely related to *Calamocichla* than to the other *Chloropeta* species (Grant and Mackworth-Praed, 1940). Several additional taxa have been the subject of taxonomic debates. For example, *A. dumetorum* has been suggested to be a subspecies of *A. scirpaceus* (Devillers and Dowsett-Lemaire, 1978; Dowsett-Lemaire and Dowsett, 1987; Hall and Moreau, 1970), or a close relative of *A. baeticatus* (Fry et al., 1974; Wilkinson and Aidley, 1983), while molecular evidence refuted the latter hypothesis (Helbig and Seibold, 1999; Leisler et al., 1997). *A. tangorum* has been treated as a subspecies of *A. bistrigiceps* (Sibley and Monroe, 1990; Watson et al., 1986; Williamson, 1968; Wolters, 1982) or *A. agricola* (Alström et al., 1991; Sibley and Monroe, 1993; Vaurie, 1959), but based on molecular data (Helbig and Seibold, 1999; Leisler et al., 1997) is now generally recognized as a separate species (del Hoyo et al., 2006; Dickinson, 2003).

The aim of this study was to clarify the relationships within the family Acrocephalidae by a combination of nuclear and mitochondrial genetic markers, targeting especially deeper nodes. Based on an alignment of a total of ~2.9 kbp, comprising the entire mitochondrial cytochrome *b* (cyt *b*), and the complete nuclear ornithine decarboxylase introns 6 and 7 (ODC), myoglobin intron 2 (myo) and lactate dehydrogenase intron 3 (LDH) for 34 species currently placed in the genera *Acrocephalus*, *Hippolais* and *Chloropeta*, this is the most comprehensive molecular analysis of this group so far. Previous analyses have been based exclusively on mitochondrial data. We also included another nine *Acrocephalus* species, for which only cyt *b* was available, and one *Nesillas* species, for which only ODC and myo were available. *Acrocephalus* warblers, in particular, have been the subject of many different types of evolutionary studies (see summary in Leisler et al., 1997), and future research will benefit greatly from the more robust phylogeny provided here.

2. Materials and methods

2.1. Study group

We obtained blood, feathers or muscle tissue from a total of 34 species of *Acrocephalus*, *Hippolais* and *Chloropeta*, as well as *Bradypterus baboecala*, *Megalurus palustris* and *Locustella lanceolata* as outgroup (Appendix A). Cyt *b* sequences from nine additional species of *Acrocephalus*, and ODC and myo from *Nesillas typica* were downloaded from GenBank (Appendix A). In all, we got sequence data from all species of *Acrocephalus* except one Eurasian (*A. sorghophilus*), one Indian Ocean (*A. rodericanus*), and four Pacific island ones (*A. luscinius*, *A. rehsei*, *A. syrinx*, *A. vaughani*); all species of *Hippolais*; and all species of *Chloropeta* (taxonomy following Cibois et al., 2007; del Hoyo et al., 2006; Parkin et al., 2004). Only extant species were used in this study. Generic/subgeneric names follow Leisler et al. (1997) and Helbig and Seibold (1999) (Table 1).

Table 2
PCR and sequencing primers (5′–3′) used for cytochrome *b* (cyt *b*), lactate dehydrogenase (LDH), ornithine decarboxylase (ODC) and myoglobin (myo).

PCR Primer	Sequence	Reference
Cyt b		
tRNALeu	ayc ttg gtg caa atc caa gt	Helbig et al. (2005)
mtFNP	ggy tta caa gac caa tgt tt	
mtF24	ttt ggt tta caa gac caa tgt ttt	
L 14841	cca tcc aac atc tca gca tga tga aa	Kocher et al. (1989)
H 15915	aac tgc agt cat ctc cgg ttt aca aga c	Edwards and Wilson (1990)
LDH		
b1	gga aga caa act aaa agg aga aat gat gga	Helbig et al. (2005)(modified)
b4	ggg ctg tat ttn acr atc tga gg	Helbig et al. (2005)
P5	gct tgc tct ggt tga ygt tat gg	
P6	cac att cct ctg cac yag gtt gag	
ODC		
OD6	gac tcc aaa gca gtt tgt cgt ctc agt gt	Allen and Omland (2003)
OD8r	tct tca gag cca ggg aag cca cca cca at	Allen and Omland (2003)
Myo		
Myo 2	gcc acc aag cac aag atc cc	Slade et al. (1993)
Myo 3F	ttc agc aag gac ctt gat aat gac tt	Heslewood et al. (1998)
Sequencing Primers		
Cyt b		
mt B-NP	cct cag aag gat aty tgk cct ca	Helbig et al. (2005)
mt B-Acro	cct cag aat gat aty tgt cct ca	
mt E-Acro	ggg ttg gck ggy gtg aar ttt tc	
mt E-Syl	ggg ttr gck ggb gtg aar ttt tc	
mt D-Syl	gac tgy gay aaa aty cca ttc ca	
mt F21	ggt tta caa gac caa tgt ttt	
LDH		
P6	att cct ctg cac yag gtt gag	
b1	cta aaa gga gaa atg atg ga	Helbig et al. (2005)
ODC	Same as PCR primer	
Myo	Same as PCR primer	

2.2. DNA extraction, amplification, sequencing and assembly

Blood samples were stored in ethanol (>99.8%), saturated soap lye or NaF-EDTA buffer, and tissue samples were stored in ethanol (>99.8%), all at −18 °C. DNA was extracted according to Miller et al. (1988) with slight modifications or using a QIAamp® DNA MiniKit (50) following the manufacturer's protocol. The following loci were sequenced: the entire mitochondrial cytochrome *b* gene (cyt *b*; 1143 bp), the nuclear ornithine decarboxylase exon 6 (partial), intron 6, exon 7, intron 7 and exon 8 (partial) (ODC; in total 740 bp), the entire nuclear myoglobin intron 2 (myo; 703 bp), and the complete nuclear lactate dehydrogenase (LDH) intron 3 (520 bp). For sequences of PCR and sequencing primers for cyt *b*, LDH, ODC and myo see Table 2. The cyt *b* gene was amplified including flanking parts to reduce the risk of amplifying nuclear copies (numts) (Sorenson and Quinn, 1998). PCRs were made up by single components or with Ready-To-GoTM PCR beads from GE Healthcare. PCR products were cleaned with ExoSap IT and products from cycle sequencing were cleaned with DyeEx 96Plate from Quiagen (only when the ABI sequencer was used). Sequencing was done on a LiCor DNA Sequencer Long READIR 4200 or on an ABI 3130xl Genetic Analyzer. Sequences were assembled manually in BioEdit (Hall, 1999) or with the Staden Package (Bonfield et al., 1995). GenBank accession numbers are given in the Appendix A.

2.3. Phylogenetic analysis

Sequences were aligned by eye in BioEdit (Hall, 1999) or with the online program T-Coffee (Notredame et al., 2000). Own incomplete sequences from cyt *b* were complemented with GenBank sequences of the same taxon to full length (see Appendix A for Accession numbers of own and GenBank sequences). As this is not a population study or an attempt to identify different subspecies, this should have no effect on the topology. Base compositions of the four different genetic markers were tested for nucleotide bias using a χ^2 test of homogeneity across taxa implemented in PAUP* 4.0b10 (Swofford, 2003). Phylogenetic analyses were performed by Bayesian inference (BI) using MrBayes 3.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), maximum likelihood (ML) with Treefinder version October 2008 (Jobb et al., 2004; Jobb, 2008), and maximum parsimony (MP) in PAUP* (Swofford, 2003). MrModeltest (Nylander, 2004) was used in conjunction with PAUP* to estimate appropriate nucleotide substitution models for implementation in MrBayes, based on the Akaike Information

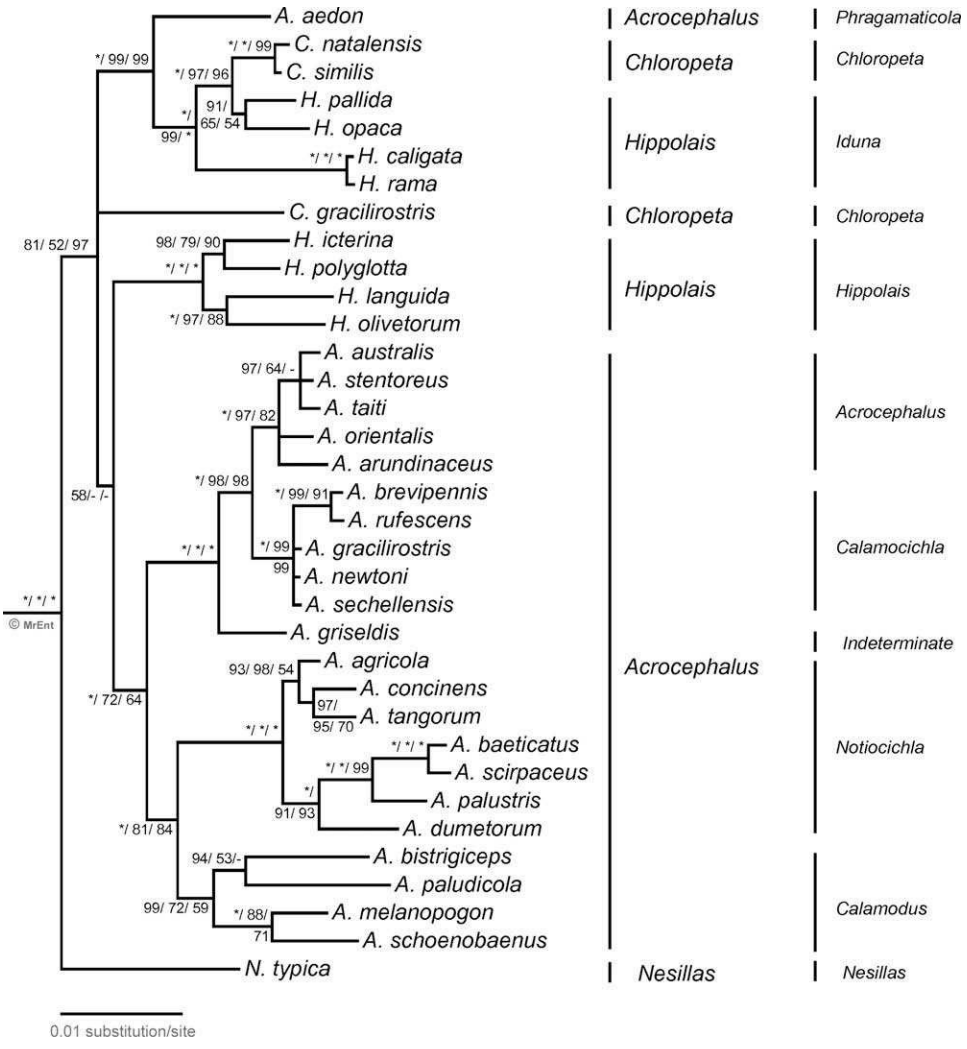

Fig. 1. Tree based on concatenated nuclear ODC, LDH and myoglobin intron sequences, analysed by Bayesian inference. Support values are shown in the following order: Bayesian posterior probabilities (⩾50%, 85,000 trees)/maximum likelihood bootstrap (⩾50%, 2000 replicates)/parsimony bootstrap (⩾50%, 2000 replicates). Asterisks indicate posterior probability = 1.00 or bootstrap = 100%. The outgroup has been pruned from the tree. Names on the right indicate generic names as traditionally defined (left), and subgenera as defined by Helbig and Seibold (1999) on the right.

Criterion (AIC; Akaike, 1973). For the ML analyses the internal model proposer of Treefinder was used to estimate nucleotide substitution models based on the AIC. Cyt *b* was not partitioned by codon position, because the variation of the second position was too low to meaningfully fit separate models. The same held for the exon of ODC, which contained only three parsimony informative sites, which provided too little information to meaningfully fit a separate model. The ODC exon was therefore excluded from the analyses in MrBayes and Treefinder, because exons and introns evolve under different conditions and therefore should not be subsumed under a single substitution model. The chosen models by MrModeltest were GTR + I + G for cyt *b* and HKY + I, HKY + G and GTR for the introns of ODC, LDH and myo, respectively. Convergence of parameters in Bayesian analysis was monitored using the program Tracer v. 1.4 (Drummond and Rambaut, 2007). In addition, the standard deviation of split frequencies was checked to ascertain optimal convergence of the chains below 0.01. Every 100th of a total number of generations ranging between 5 and 11 million was sampled, the burnin amounted to 25%. The models for ML bootstrapping (2000 replicates) chosen by Treefinder were GTR + G for cyt *b*, HKY for ODC, TVM + G for LDH and TVM for myo. MP bootstrapping was performed under the following settings: heuristic search, starting trees obtained by random stepwise addition, 2000 replicates and TBR branch swapping. Five data sets were analysed: cyt *b*; the nuclear introns of ODC, LDH, myo separately and combined; and all four concatenated partitions. Indels were treated as missing data in BI and ML. In order to incorporate the information of indels appropriately in the MP analyses, modified complex indel coding was applied in SeqState (Müller, 2005, 2006). Posterior probabilities $\geqslant 0.95$ and bootstrap values $\geqslant 0.85$ were regarded as high support following Erixon et al. (2003).

For testing the monophyly of different groups, constraint trees were compared against the optimal tree employing approximately unbiased tests (Shimodaira, 2002) with 10,000 replicates in a ML framework in Treefinder. The monophyly of indels was also tested in a MP framework by Templeton tests using PAUP*.

Bradypterus baboecala, *Megalurus palustris* and *Locustella lanceolata* (Megaluridae) were chosen as outgroup, as they are part of the sister clade to Acrocephalidae according to Alström et al. (2006) and Johansson et al. (2008).

3. Results

3.1. Phylogeny

Including indels, the concatenated cyt *b*, ODC, LDH and myo alignment comprised 2919 positions. Base compositions of each

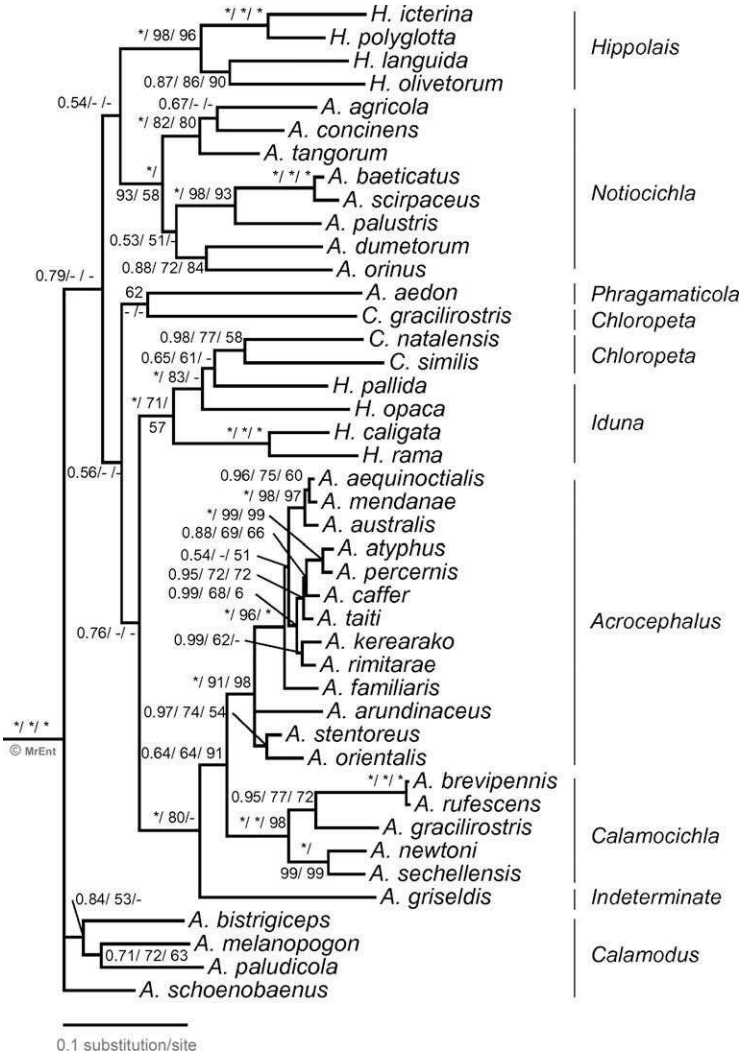

Fig. 2. Tree based on the mitochondrial cytochrome *b* gene, analysed by Bayesian inference. Support values marked as in Fig. 1. Names on the right indicate subgenera as defined by Helbig and Seibold (1999).

partition and across the entire data set were homogeneous. Cyt *b* had 414 (36%) parsimony informative sites, ODC 73 (10%), LDH 66 (13%) and myo 57 (8%), totalling 20% across the complete alignment.

All four loci could be safely combined, since there was no significant conflict between the analyses of the individual loci. Only three trees are shown here: the combined nuclear dataset (ODC, LDH and myo; hereafter called nuclear dataset) in Fig. 1; the mitochondrial cyt *b* in Fig. 2; and the overall combined cyt *b*, ODC, LDH and myo tree (hereafter called complete dataset) in Figs. 3 and 4.

The ingroup, Acrocephalidae, was recovered with strong support in all analyses, and the outgroup was therefore pruned from the trees in the figures. As expected, the mitochondrial cyt *b* was more variable than the nuclear loci, and therefore better depicted species relationships, whereas the nuclear loci were much more conservative and better at resolving deeper nodes. All shown trees from the separate and combined partitions (Figs. 1–4) recovered basically the same six major clades, regardless of method: (1) four species of *Hippolais* (*Iduna*), *Acrocephalus aedon, Chloropeta natalensis* and *C. similis* (*A. aedon* excluded from this clade in cyt *b* tree); (2) four other *Hippolais* (s.s.); (3) 5–13 Eurasian to Polynesian large *Acrocephalus* (i.e. *Acrocephalus* s.s.; number depending on analysis – for eight species only cyt *b* available); (4) the five Afrotropical/Malagasy *Acrocephalus* (*Calamocichla*); (5) 6–7 Eurasian and one Afrotropical small unstreaked *Acrocephalus* (*Notiocichla*; for one species only cyt *b* available); and (6) four small Eurasian *Acrocephalus*, all except one of which are streaked (*Calamodus*; *A. schoenobaenus* excluded from clade in cyt *b* tree).

As is evident from the above, none of the three genera was recovered as monophyletic. *A. aedon* was sister to a clade containing four *Hippolais* (*Iduna*) and two *Chloropeta* in the nuclear (Fig. 1) and complete (Figs. 3 and 4) dataset trees. This relationship was

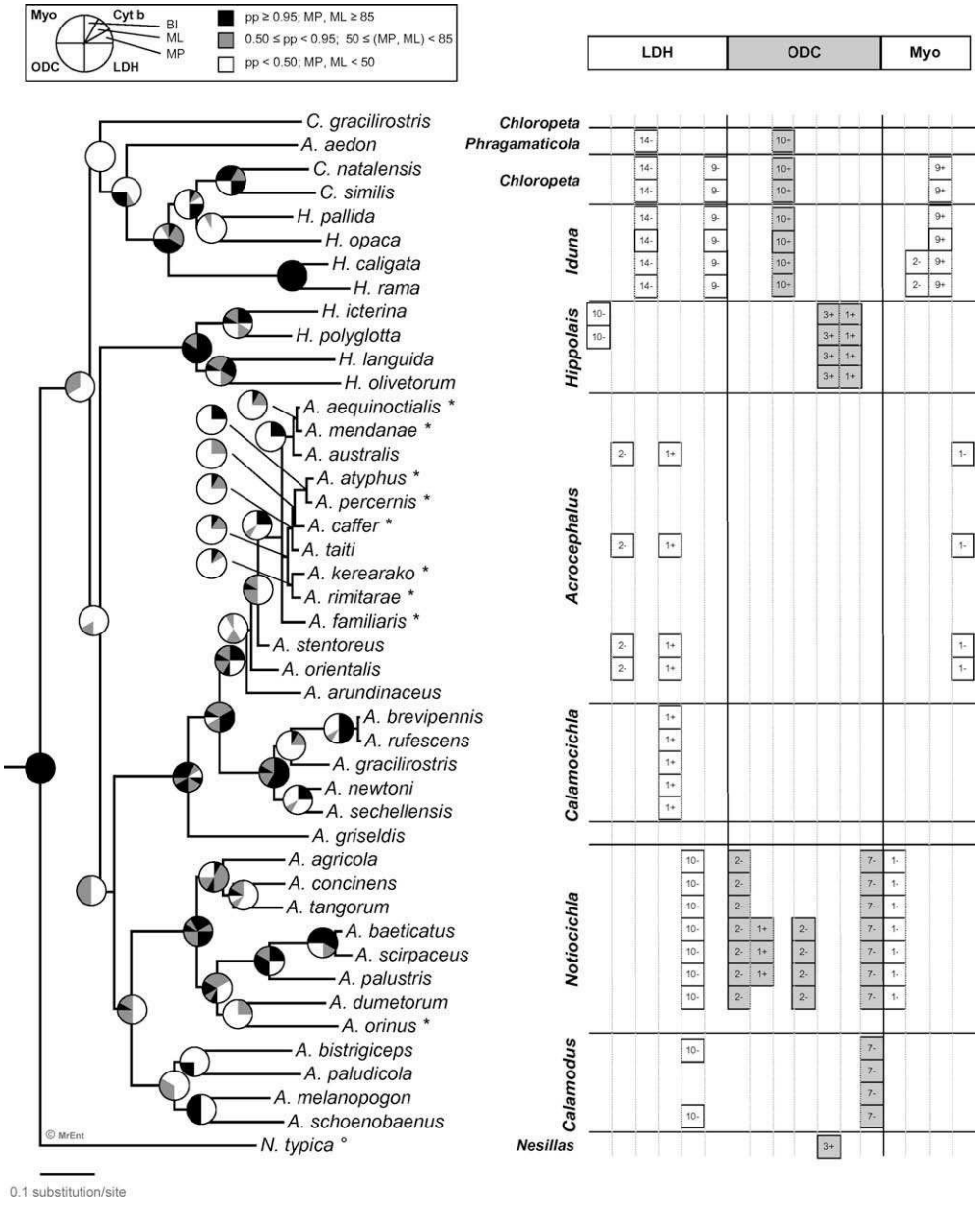

Fig. 3. Tree based on the complete concatenated dataset (cytochrome *b*, LDH, ODC and myoglobin), analysed by Bayesian inference. Pie charts indicate support in the four single-locus analyses (see explanation in upper left corner of figure). Squares on the right indicate indels shared among taxa in the alignments of the non-coding regions. Numbers in squares indicate size of indels: + means insertion and deletion. Names in bold refer to subgenera as defined by Helbig and Seibold (1999). Taxa marked with an asterisk (*): only cyt *b* was available. Taxon marked with a circle (°): only ODC and myo were available.

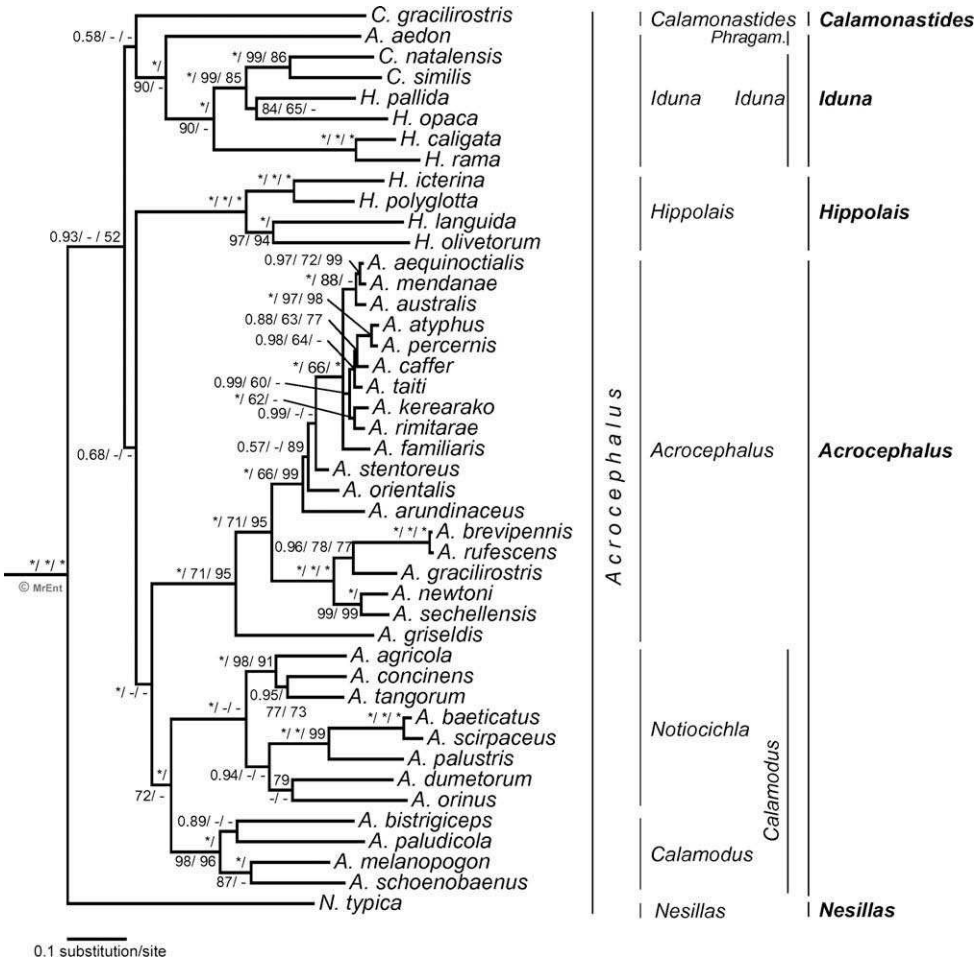

Fig. 4. Tree based on the complete concatenated dataset (cytochrome *b*, LDH, ODC and myoglobin), analysed by Bayesian inference. Support values marked as in Fig. 1. On the right, different possibilities for a revised taxonomy are shown. Taxonomy favoured by the authors in bold.

further supported by a 14 bp deletion in the LDH alignment and a 10 bp insertion in the ODC alignment. The position of *A. aedon* was effectively unresolved in the cyt *b* tree (Fig. 2). This *Iduna/Chloropeta* clade was well supported in all shown analyses, and was also corroborated by a 9 bp LDH deletion and a 9 bp myo insertion. *Hippolais* s.s. was also well supported in every analysis as well as by two short insertions in the ODC alignment. It was consistently found separated from the other clade containing *Hippolais* (*Iduna*), although its exact position varied among the analyses and never received strong support. *Nesillas typica* was recovered as sister to all other taxa in the nuclear and combined data set (only ODC and myo available). *Chloropeta gracilirostris* held an enigmatic position in all analyses.

Taxa from the genus *Acrocephalus* split up into four main clades, corresponding to *Acrocephalus* s.s., *Calamocichla*, *Notiocichla* and *Calamodus*, mostly well supported, also by indels (Fig. 4). *Notiocichla* and *Calamodus* were supported as sister groups by a 7 bp deletion in the ODC alignment. The first one was further supported by a 2 bp ODC deletion and a 1 bp deletion in the myo alignment. *Acrocephalus* (s.s.) and *Calamocichla* formed a clade in all trees shown here, with high support in the nuclear (Fig. 1) and complete (Fig. 4) dataset analyses, but poor support in the cyt *b* tree (Fig. 2). All species in these two clades except *A. arundinaceus* also shared a 1 bp insertion in the LDH alignment, and all species in the *Acrocephalus* s.s. clade except *A. arundinaceus* had two unique deletions, 2 bp in the LDH alignment and 1 bp in the myo alignment (Fig. 4). *A. griseldis* was recovered as sister to *Acrocephalus* s.s. and *Calamocichla* in all shown analyses (except cyt *b* MP analysis).

At a lower level, *Calamocichla* separated into two sister groups with good support in the complete (Fig. 4) and cyt *b* analyses (Fig. 2): one clade containing the Malagasy *A. newtoni* and Seychelles *A. sechellensis* and another clade with the mainly eastern/southern Afrotropical *A. gracilirostris* as sister to the mostly central/western Afrotropical *A. rufescens* and the Cape Verde endemic *A. brevipennis*. In the nuclear tree, the two latter were sisters, while the three others were in an unresolved position in this clade. In *Acrocephalus* (s.s.), all Australian and Polynesian species (*A. aequinoctialis, A. mendanae, A. australis*, *A. atyphus, A. percernis*, *A. caffer*, *A. taiti*, *A. kerearako, A. rimitarae, A. familiaris*) formed a well supported clade in the cyt *b* (Fig. 2) and complete data tree (Fig. 4) (though in the latter, this was based exclusively on cyt *b* data for all except *A. australis* and *A. taiti*). In both these trees, this clade was separated from the south Asian *A. stentoreus*, east Asian *A. orientalis*, and Palearctic *A. arundinaceus*. In the nuclear tree, *A. stentoreus* instead formed a clade with *A. australis* and *A. taiti*, separated from *A. orientalis* and *A. arundinaceus*, with good BI support, but poor ML and MP support. This and a sister relationship between *A. australis* and *A. orientalis*, only supported by BI in the LDH analysis (not shown here) were the only cases of incongruence between different data sets that were well supported in any analysis.

Within *Notiocichla*, *A. agricola*, *A. concinens* and *A. tangorum* formed a mostly well supported clade. The relationship between these taxa was best resolved in the nuclear analysis. The clade

consisting of *A. palustris* as sister taxon to *A. baeticatus* and *A. scirpaceus* was well supported in all shown trees, as was the sister relationship between the two latter; this clade was further supported by a 1 bp insertion in the ODC alignment (Fig. 3). *A. dumetorum* was supported as sister taxon to this constellation in the complete and nuclear data set and by a 2 bp deletion in ODC. *A. orinus* was recovered with poor support as sister taxon to *A. dumetorum* in the complete and cyt *b* data set (no nuclear data available). The relationships within *Calamodus* were not unambiguously resolved, and especially in the cyt *b* tree (Fig. 2) they were best regarded as unresolved. *Iduna* split into two sister groups with high support: *C. natalensis*, *C. similis*, *H. pallida* and *H. opaca* formed one clade, while *H. caligata* and *H. rama* formed the other. The relationships within the first clade varied between the combined/nuclear and cyt *b* trees, with a *C. natalensis*/C. *similis* sister relationship on one side and *H. pallida*/*H. opaca* on the other (only nuclear and combined trees). The *Hippolais* (s.s.) clade was well supported in all shown analyses, with *H. languida*/*H. olivetorum* and *H. polyglotta*/*H. icterina* as sisters.

Although the monophyly of *Acrocephalus* sensu lato (s.l.) was not supported, it could not be rejected by the approximately unbiased test comparing a tree constraining all *Acrocephalus* to be monophyletic and the unconstrained tree ($p = 0.357$). In contrast, the monophyly of the genera *Hippolais* and *Chloropeta* was significantly rejected ($p < 0.001$).

3.2. Indels

As presented above, the aligned introns of ODC and LDH contained phylogenetically informative indels, which were highly correlated with the other phylogenetic signal. However, two indels were apparently homoplastic. A 10 bp LDH deletion of *Notiocichla* (Fig. 3) was also shared by two members of its sister clade, *A. bistrigiceps* and *A. schoenobaenus*. In addition, a 1 bp insertion defined *Acrocephalus* s.s. and *Calamocichla* except *A. arundinaceus*. In order to test whether such complex evolutionary events can indeed be homoplastic, a tree constraining all indels to define monophyletic groups was compared to the optimal, unconstrained reconstructions in both ML and MP frameworks. According to the approximately unbiased test, the likelihood of the constraint tree was significantly lower ($p < 0.001$). Under the criterion of MP, the constraint tree was significantly longer (Templeton test, $p < 0.001$).

4. Discussion

4.1. Phylogeny

The present study provides an overall better resolved hypothesis of relationships than the mitochondrial studies of Leisler et al. (1997) and Helbig and Seibold (1999). Moreover, importantly, the addition of independent data in the form of three unlinked nuclear loci results in improved support values and an overall better corroborated hypothesis. However, deep internal nodes are still poorly resolved. Most of these are very short, indicating rapid radiations (as suggested by Helbig and Seibold, 1999). If that is indeed the case, addition of further loci may still fail to fully resolve the phylogeny (cf. e.g. Belfiore et al., 2008; Wiens et al., 2008). In general, the results of Leisler et al. (1997) and Helbig and Seibold (1999) are largely confirmed, recovering the same major clades. Although there is strong support that *A. aedon* belongs in a clade with *C. natalensis, C. similis* and four species of *Hippolais* (*Iduna*), and accordingly that the genus *Acrocephalus* as usually circumscribed is non-monophyletic, the approximately unbiased test could neither support nor reject the monophyly of *Acrocephalus*. However, the inclusion of additional species, more data and the more advanced approaches of ML and BI did help to clarify a number of details. Both *Hippolais* s.l. and *Chloropeta* are conclusively shown to be non-monophyletic. The monophyly of the genus *Chloropeta* has been questioned before. Grant and Mackworth-Praed (1940) proposed that *C. gracilirostris* should be placed in the monotypic genus *Calamonastides*, because they saw closer similarities to *Calamocichla* than to other *Chloropeta* species. A close relationship between *C. gracilirostris* and *Calamocichla* is not supported by our analysis, although the precise position of *C. gracilirostris* remains obscure.

The position of *C. natalensis* and *C. similis* within the *Iduna* clade, and the sister relationship between this clade and *A. aedon* are unexpected. *C. natalensis* and *C. similis* have not been included in previous studies, and *A. aedon* was only included by Helbig and Seibold (1999), who were unable to resolve its relationships. The inclusion of *C. natalensis* and *C. similis* in the *Iduna* clade is well corroborated by our data. In terms of size, both fit well into the *Iduna* assemblage, while its different coloration and bill shape (del Hoyo et al., 2006) are probably autapomorphic adaptations. In addition, the habitat requirements agree well between these *Chloropeta* species and the *Iduna* species (del Hoyo et al., 2006). Also from a geographical point of view, the close relationship between *C. natalensis*/*C. similis* and especially *H. opaca* and *H. pallida* makes sense; the latter has subspecies which are sedentary in Africa, like *C. natalensis* and *C. similis* (del Hoyo et al., 2006). The sister relationship between *C. natalensis* and *C. similis* is well supported by our data. However, whether *H. pallida* is most closely related to *H. opaca* or to *C. natalensis*/*C. similis* is uncertain, as the mitochondrial and nuclear data disagree (none strongly supported). This is interesting, since until recently *H. pallida* and *H. opaca* were treated as conspecific (Ottosson et al., 2005; Parkin et al., 2004), while *C. natalensis* and *C. similis* are usually placed in a different genus (e.g. del Hoyo et al., 2006; Dickinson, 2003; Sibley and Monroe, 1990; Wolters, 1982). In contrast, the sister relationship between *H. caligata* and *H. rama* are well supported in all analyses. The position of *A. aedon* is less firmly anchored, as it is only resolved in the single-locus analyses of ODC. However, the 10 bp ODC insertion and14 bp LDH deletion provide additional support for this reconstruction. *A. aedon* is considerably larger and differently shaped than the remaining species, but its egg coloration is stated to be similar to that of *Hippolais* and *C. natalensis* but different from *Acrocephalus* (Schönwetter, 1979). In addition, *A. aedon* has a peculiar egg shell graining similar to *Hippolais* and *Nesillas* (Schönwetter, 1979). Its preferred breeding habitat in drier habitats with thickets is also in concordance with the *Iduna* species (del Hoyo et al., 2006). Also its nest construction and song are more like *Hippolais* than other *Acrocephalus* species (Marova et al., 2005).

The *Hippolais* s.s. clade and the relationships within this clade are strongly supported, and although none of the analyses suggest a sister relationship with the clade comprising the other species traditionally placed in *Hippolais* (*Iduna*), the possibility of a sister relationship between these two is not strongly rejected by our data.

Acrocephalus s.s., comprising the large Eurasian, Polynesian and Australian species, is well supported. The affinities of the eight Polynesian species *A. aequinoctialis*, *A. mendanae*, *A. atyphus*, *A. percernis*, *A. caffer*, *A. kerearako*, *A. rimitarae* and *A. familiaris* to the others in this clade have not been shown before, although this has generally been assumed to be the case, since they have been placed close to each other in linear taxonomies (e.g. del Hoyo et al., 2006; Dickinson, 2003; Sibley and Monroe, 1990; Watson et al., 1986). Recent studies (Fleischer et al., 2007; Cibois et al., 2007) dealt only with the relationships within the Polynesian species. However, within *Acrocephalus* s.s., the relationships are generally not very well supported (for nearly half of the species only cyt *b* was available), and there is some conflict between

the mitochondrial and nuclear data with respect to the position of *A. stentoreus*. More data will be needed to elucidate these relationships.

The sister relationship between *Acrocephalus* s.s. and *Calamocichla* is well corroborated, and unites all of the large *Acrocephalus* (s.l.) species except *A. aedon* (see above) and *A. griseldis* (see below); two small species (*A. brevipennis* and *A. sechellensis*) are also included. The African/Malagasy/Indian Ocean radiation (*Calamocichla*) is also strongly supported, as are the internal relationships within this clade (except for the position of *A. gracilirostris*). The position of *A. griseldis* as sister to the clade comprising *Acrocephalus* s.s. and *Calamocichla*, as suggested by Leisler et al. (1997) and Helbig and Seibold (1999) based on cyt *b*, is strongly corroborated here. In agreement with nearly all of the species in *Acrocephalus* s.s. and *Calamocichla*, *A. griseldis* is a large species (del Hoyo et al., 2006).

The clade referred to as *Notiocichla* by Leisler (1997) and Helbig and Seibold (1999) is strongly supported. Of the two main subclades in this clade, the one comprising *A. agricola*, *A. concinens* and *A. tangorum* is well supported. The other one, comprising *A. baeticatus*, *A. scirpaceus*, *A. palustris*, *A. dumetorum* and *A. orinus*, is well supported only in the BI analyses of the complete dataset. However, in the nuclear tree this clade, excluding *A. orinus*, for which we have no nuclear sequences, is strongly supported. Moreover, this clade is recovered by three of the four loci and further supported by an ODC deletion, so we believe that it is well corroborated. Within the first-mentioned subclade, *A. tangorum* and *A. concinens* are found to be sisters in the complete tree, but the support for this is not unanimously strong, and apparently rests mainly on one locus (myo). Within the second subclade in *Notiocichla*, *A. dumetorum* is inferred to be sister to *A. orinus* (see below), which together form a sister clade to the *A. baeticatus/A. scirpaceus/A. palustris* clade. Leisler et al. (1997) came to the same conclusion based on cyt *b* sequence data (although they did not include *A. orinus*), while Haffer's (1991) suggestion that *A. dumetorum* is sister to *A. palustris* is rejected by our results. Also the suggested close relationship between *A. dumetorum* and *A. baeticatus* (Fry et al., 1974; Wilkinson and Aidley, 1983) is refuted (in agreement with Leisler et al., 1997 and Helbig and Seibold, 1999). The sister relationship between *A. dumetorum* and the recently rediscovered Central Asian *A. orinus* (Round et al., 2007; Svensson et al., 2008) is not strongly supported by our data, and since it is only based on cyt *b* it must be considered tentative (our attempts to amplify other loci from one fresh sample and several museum specimens have failed). Surprisingly, Round et al. (2007) received stronger BI support (0.98) for this relationship based on the same cyt *b* sequence.

The clade, which was referred to as *Calamodus* by Leisler et al. (1997) and Helbig and Seibold (1999) is well supported. This agrees with morphological and vocal characteristics (Leisler et al., 1997). We hypothesize that the little-known east Asian *A. sorghophilus* belongs in this clade, probably as sister to *A. bistrigiceps*. The relationships within *Calamodus* are somewhat uncertain, since there is incongruence between different loci, and since the topology in the combined mitochondrial and nuclear tree is not well supported by all methods. The sister relationship between *Notiocichla* and *Calamodus* is not unanimously strongly supported by the data. However, it makes sense from a morphological point of view (Leisler et al., 1997).

4.2. Indels

Homoplasy is apparently an important and confounding issue in acrocephalid systematics. Leisler et al. (1997) were the first ones to demonstrate discrepancies between relationships inferred based on genetic data and morphology and thus the traditional classification. The splitting of *Chloropeta* and the position of *A. aedon* further complicate the picture. In addition, we found quite complex cases of convergence also at the genetic level. *A. bistrigiceps* and *A. schoenobaenus* have independently lost the same 10 bp in the LDH intron as the species in *Notiocichla*. This event may even have occurred in parallel in *A. bistrigiceps* and *A. schoenobaenus*, but the insufficiently supported relationships within this clade do not permit a definite conclusion. Conversely, *A. paludicola* and *A. melanopogon* could have regained the nucleotides. However, we consider this less likely, because the sequence of these 10 bp was practically identical to the homologous 10 bp in all other studied species. The convergent loss of 10 bp indicates that this section of the intron is under relaxed functional constraints. The most parsimonious explanation for the lack of a 1 bp insertion in *A. arundinaceus* defining *Acrocephalus* s.s. and *Calamocichla* is a secondary loss, since this insertion is found in all other species in these clades.

5. Taxonomic consequences

Our results indicate that none of the existing classifications (Table 1) can be maintained. We see two main alternatives: (1) *Hippolais* Conrad, 1827 and *Chloropeta* A. Smith, 1847 are synonymized with *Acrocephalus* J.A. and J.F. Naumann, 1811, or (2) several smaller genera are recognized with various possibilities (Fig. 4). The first alternative would have the advantage of uniting all taxa in the very well corroborated most-inclusive clade into a single genus, which would comprise c. 53 species. However, this would have the disadvantage of creating two homonyms: *A. gracilirostris* (Hartlaub, 1864) (the current *A. gracilirostris*) and *A. gracilirostris* (Ogilvie-Grant, 1906) (presently *Chloropeta gracilirostris*). The former name has priority, so the latter taxon would have to be given a new name. *Nesillas* could perhaps be included in this expanded *Acrocephalus*, although this seems premature considering that only one species and two loci have been examined.

The main advantage of the second alternative is that it would provide more phylogenetic information than the first one, but there are different possibilities of which we outline two. Splitting rigorously, the species in the clade comprising *Iduna*, *C. natalensis, C. similis* and *A. aedon* are united in the genus *Iduna* Keyserling and Blasius, 1840, which has priority over *Chloropeta* A. Smith, 1847 (type species *C. natalensis*) and *Phragamaticola* Jerdon, 1845 (type species *P. aedon*). It would also be possible to exclude the taxon *aedon* from this genus and place it in a monotypic *Phragamaticola* on the basis that its inclusion in this clade is based on comparatively little genetic evidence, and hence not very well corroborated. *Chloropeta gracilirostris*, whose position in the tree is uncertain, is placed in the monotypic genus *Calamonastides* Grant and Mackworth-Praed, 1940; in this case, no new name would have to be proposed for this taxon. The generic name *Chloropeta* is not available for this taxon, since the type species of that genus is *C. natalensis*, which is now included in *Iduna*. The name *Hippolais* Conrad, 1827 is restricted to the well supported *Hippolais* s.s. clade, which includes the type species of this genus (*H. polyglotta*). *Acrocephalus* s.s., *Calamocichla* and *A. griseldis*, which together form a strongly supported clade, could be treated as *Acrocephalus* (s.s.). *Notiocichla* and *Calamodus* could be united under the name *Calamodus* Kaup, 1829, which has priority over *Notiocichla* Oberholser, 1905.

A more conservative classification would be to leave the genus *Acrocephalus* as it is currently established, with the exception of *A. aedon*, and retaining *Hippolais* for the *Hippolais* s.s. clade. The remaining *Hippolais* species (*Iduna* clade), together with *C. natalensis*, *C. similis* and *A. aedon* are united in *Iduna*. *C. gracilirostris* can not

be associated with certainty to any of the clades and is therefore appropriately placed in the monotypic genus *Calamonastides*. Nothing should be changed regarding the genus *Nesillas*, as only one species was included in the present study, and it seems to be in a sister position to all other Acrocephalidae.

We favor the last alternative, since this involves the fewest changes, and since we prefer larger genera to smaller ones.

Considering the weak resolution of deeper nodes, this proposal is probably not the final call towards a more natural classification of Acrocephalidae. It is certainly desirable to collect more information both in terms of sequence data as well as taxon sampling in order to better resolve the relationships of the larger clades and improve our understanding of the evolution of the family in time and space.

Acknowledgments

SF is grateful to the late A.J. Helbig as initial supervisor and for initiating this study. We are grateful to A. Berruti, L. DeBrooke, D. Brown, J. Dierschke, W. Forstmeyer, D. Franz, K. Gauger, B. Giessing, G. Halmos, A.J. Helbig, J.B. Kristensen (Zoological Museum Copenhagen, Denmark), B. Leisler, T. Michaelson, J. Morgan, G. Nikolaus, D.J. Pearson, S. Rumsey, P. Round, V. Salewski, H. Shirihai and F. Wong for providing samples. For helpful comments on the taxonomy we thank Edward Dickinson. We are also thankful for comments from two anonymous reviewers on this manuscript.

This study was partially financed by a European Union Synthesys grant (SE-TAF-2992).

Appendix A. Studied taxa in alphabetical order. Locality, GenBank accession number and sample number are given. * refers to incomplete sequences from cytochrome *b*, which were complemented with GenBank sequences. In parenthesis numbers of basepairs (bp) from GenBank. Taxonomy of subspecies follows Dickinson (2003).

Taxon	Locality	Gene	Sample No.	GenBank No.	Complemented * with GenBank No.
Acrocephalus aedon	China, Hebei	Cyt *b*	B 0650	FJ883020	AJ004778 (414 bp)
		LDH	B 0926	FJ883054	
		ODC	B 0650	FJ883126	
		Myo	B 0926	FJ883104	
Acrocephalus agricola agricola	Kazakhstan, Lake Kurgaljino	Cyt *b*	B 0917	FJ883021	AJ004775 (783 bp)
	Kazakhstan	LDH	B 0955	FJ883055	
		ODC	B 0934	FJ883127	
		Myo	B 0934	FJ883108	
Acrocephalus aequinoctialis aequinoctialis	Kiribati Is.	Cyt *b*		EF156278	
Acrocephalus arundinaceus arundinaceus	Austria	Cyt *b*	B 0647	FJ883022	AJ004784 (250 bp)
		LDH	B 0647	FJ883056	
		ODC	B 0648	FJ883128	
		Myo	B 0647	FJ883098	
Acrocephalus atyphus flavidus	Tuamotu Is, Napuka	Cyt *b*		EF156297	
Acrocephalus australis carterae/gouldi	West Australia	Cyt *b*	B 0641	FJ883023	AJ004786 (771 bp)
		LDH	B 0641	FJ883057	
		ODC	B 0641	FJ883129	
		Myo	B 0641	FJ883097	
Acrocephalus baeticatus hallae	South Africa	Cyt *b*	B 0921	FJ883024	
		LDH	B 0921	FJ883058	
		ODC	B 0921	FJ883130	
		Myo	B 0921	FJ883103	
Acrocephalus bistrigiceps	China, Hebei	Cyt *b*	B 0945	FJ883025	AJ004766 (413 bp)
		LDH	B 0945	FJ883059	
		ODC	B 0945	FJ883131	
		Myo	B 0945	FJ883109	
Acrocephalus brevipennis	Cape Verde	Cyt *b*	B 0651	FJ883026	
		LDH	B 0651	FJ883060	
		ODC	B 0651	FJ883132	
		Myo	B 0651	FJ883099	
Acrocephalus caffer	Society Is	Cyt *b*		EF156308	
Acrocephalus concinens concinens	Thailand	Cyt *b*	B 0948	FJ883027	
	China, Hebei	LDH	ZMUC 01944	FJ883061	
		ODC	ZMUC 01944	FJ883133	
		Myo	ZMUC 01944	FJ883101	
Acrocephalus dumetorum	Russia, Yenessei	Cyt *b*	B 0949	FJ883028	AJ004773 (780 bp)
	E Kazakhstan	LDH	B 0574	FJ883062	
	Russia, Yenessei	ODC	B 0928	FJ883134	
		Myo	B 0928	FJ883105	
Acrocephalus familiaris kingi		Cyt *b*		EU119965	
Acrocephalus gracilirostris parvus	SE Kenya	Cyt *b*	B 0582	FJ883029	AJ004782 (409 bp)
		LDH	B 0582	FJ883063	
		ODC	B 0582	FJ883135	
		Myo	B 0582	FJ883095	
Acrocephalus griseldis	Kenya	Cyt *b*	B 0513	FJ883030	AJ004779 (275 bp)
		LDH	B 0513	FJ883064	

(*continued on next page*)

Appendix A (*continued*)

Taxon	Locality	Gene	Sample No.	GenBank No.	Complemented * with GenBank No.
		ODC	B 0513	FJ883136	
		Myo	B 0513	FJ883092	
Acrocephalus kerearako kerearako	Cook Is, Mangaia	Cyt *b*		EF156292	
Acrocephalus melanopogon melanopogon	Hungary	Cyt *b*	B 0938	FJ883032	AJ004767 (780 bp)
		LDH	B 0938	FJ883066	
Acrocephalus melanopogon mimicus	Azerbaijan	ODC	B 0530	FJ883138	
		Myo	B 0530	FJ883093	
Acrocephalus mendanae mendanae	Marquesas Is, Tahuata	Cyt *b*		EF156310	
Acrocephalus newtoni	Madagascar	Cyt *b*	B 1031	FJ883033	AJ004780 (769 bp)
		LDH	B 1032	FJ883067	
		ODC	B 1031	FJ883139	
		Myo	B 1031	FJ883123	
Acrocephalus orientalis	Hong Kong	Cyt *b*	B 0920	FJ883034	AJ004785 (770 bp)
		LDH	B 0920	FJ883068	
		ODC	B 0920	FJ883140	
		Myo	B 0920	FJ883102	
Acrocephalus orinus	Thailand	Cyt *b*		DQ681065	
Acrocephalus paludicola	Poland	Cyt *b*	B 0957	FJ883035	AJ004768 (443 bp)
		LDH	B 0957	FJ883069	
		ODC	B 0943	FJ883141	
		Myo	B 0957	FJ883112	
Acrocephalus palustris	Germany	Cyt *b*	B 0930	FJ883036	AJ004774 (776 bp)
		LDH	B 0930	FJ883070	
		ODC	B 0930	FJ883142	
		Myo	B 0930	FJ883106	
Acrocephalus percernis percernis		Cyt *b*		EF156298	
Acrocephalus rimitarae	Austral Is, Rimatara	Cyt *b*		EF156306	
Acrocephalus rufescens senegalensis	Senegal	Cyt *b*	B 1060	FJ883037	
		LDH	B 1060	FJ883071	
		ODC	B 1060	FJ883143	
		Myo	B 1060	FJ883124	
Acrocephalus schoenobaenus	Russia, Yenessei	Cyt *b*	B 0953	FJ883038	Z73475 (407 bp)
	Oman	LDH	B 0526	FJ883072	
		ODC	B 0931	FJ883144	
		Myo	B 0931	FJ883107	
Acrocephalus scirpaceus ssp.	Germany	Cyt *b*	B 0939	FJ883039	Z73483 (773 bp)
	Israel, Kenneret	LDH	B 0950	FJ883073	
		ODC	B 0950	FJ883145	
		Myo	B 0950	FJ883111	
Acrocephalus sechellensis	Seychelles	Cyt *b*	B 1030	FJ883040	AJ004781 (772 bp)
		LDH	B 1030	FJ883074	
		ODC	B 1030	FJ883146	
		Myo	B 1030	FJ883122	
Acrocephalus stentoreus levantinus	Israel, Dead Sea	Cyt *b*	B 0572	FJ883031	AJ004787 (408 bp)
		LDH	B 0572	FJ883065	
		ODC	B 0918	FJ883137	
		Myo	B 0918	FJ883100	
Acrocephalus taiti	Henderson Island	Cyt *b*	B 0639	FJ883042	
		LDH	B 0640	FJ883076	
		ODC	B 0640	FJ883148	
		Myo	B 0640	FJ883096	
Acrocephalus tangorum	Thailand	Cyt *b*	B 0947	FJ883041	AJ004777 (414 bp)
		LDH	B 0947	FJ883075	
		ODC	B 0947	FJ883147	
		Myo	B 0947	FJ883110	
Bradypterus baboecala elgonensis	Kenya	Cyt *b*	B 0706	FJ883053	
Bradypterus baboecala tongensis/transvaalensis	South Africa	LDH	B 1018	FJ883090	
Bradypterus baboecala elgonensis	Kenya	ODC	B 0706	FJ883162	
Bradypterus baboecala tongensis/transvaalensis	South Africa	Myo		DQ008525	
Chloropeta gracilirostris gracilirostris	Kenya	Cyt *b*	B 0511	FJ883043	
		LDH	B 0511	FJ883077	
		ODC	B 0960	FJ883149	
		Myo	B 0960	FJ883113	
Chloropeta natalensis natalensis		Cyt *b*		DQ008523	
Chloropeta natalensis massaica	Uganda, Ruwenzori Mts	LDH	ZMUC 123044	FJ883082	
		ODC	ZMUC 123044	FJ883150	
		Myo	ZMUC 123044	FJ883114	

Appendix A (*continued*)

Taxon	Locality	Gene	Sample No.	GenBank No.	Complemented * with GenBank No.
Chloropeta similis	Kenya	Cyt *b*	ZMUC 131329	FJ899738	
		LDH	ZMUC 131329	FJ883083	
		ODC	ZMUC 131329	FJ883159	
		Myo	ZMUC 131329	FJ883125	
Hippolais caligata	Kazakhstan	Cyt *b*	B 0967	FJ883044	AJ004793 (423 bp)
	India, Karnataka	LDH	B 0969	FJ883084	
		ODC	B 0969	FJ883151	
		Myo	B 0969	FJ883116	
Hippolais icterina	Greece	Cyt *b*	B 0509	FJ883046	DQ008479.1 (410 bp)
	Ukraine, Crimea	LDH	B 0981	FJ883078	
		ODC	B 0981	FJ883153	
		Myo	B 0981	FJ883120	
Hippolais languida	Kenya	Cyt *b*	B 0544	FJ883047	AJ004794 (410 bp)
		LDH	B 0544	FJ883079	
		ODC	B 0544	FJ883154	
		Myo	B 0544	FJ883094	
Hippolais olivetorum	Kenya	Cyt *b*	B 0541	FJ883048	AJ004795 (781 bp)
		LDH	B 0541	FJ883080	
		ODC	B 0983	FJ883155	
		Myo	B 0983	FJ883121	
Hippolais opaca	Senegal	Cyt *b*	B 0507	FJ883049	AJ004790 (432 bp)
		LDH	B 0507	FJ883086	
		ODC	B 0975	FJ883156	
		Myo	B 0975	FJ883118	
Hippolais pallida elaeica	Kenya, Ngulia	Cyt *b*	B 0537	FJ883045	AJ004791 (771 bp)
		LDH	B 0537	FJ883085	
		ODC	B 0977	FJ883152	
		Myo	B 0977	FJ883119	
Hippolais polyglotta	Ivory Coast	Cyt *b*	B 0964	FJ883050	AJ004797 (782 bp)
		LDH	B 0964	FJ883081	
		ODC	B 0964	FJ883157	
		Myo	B 0964	FJ883115	
Hippolais rama	Turkmenistan	Cyt *b*	B 0971	FJ883051	AJ004792 (772 bp)
		LDH	B 0971	FJ883087	
		ODC	B 0971	FJ883158	
		Myo	B 0971	FJ883117	
Locustella lanceolata		Cyt *b*		DQ119525.1	
	China, Hebei	LDH	B 0761	FJ883088	
		ODC	HI000523BP	FJ883160	
		Myo		FJ883091	
Megalurus palustris forbesi	Philippines, Isabella	Cyt *b*	ZMUC 02031	FJ883052	
		LDH	ZMUC 02031	FJ883089	
		ODC	ZMUC 02031	FJ883161	
Megalurus palustris toklao	India, Punjab	Myo		DQ008529	
Nesilla typica		ODC		EU680744	
		Myo		EU680592	

References

Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

Allen, E.S., Omland, K.E., 2003. Novel intron phylogeny (ODC) supports plumage convergence in orioles (Icterus). Auk 120, 961–969.

Alström, P., Olsson, U., Round, P.D., 1991. The taxonomic status of *Acrocephalus agricola tangorum*. Forktail 6, 3–13.

Alström, P., Ericson, P.G.P., Olsson, U., Sundberg, P., 2006. Phylogeny and classification of the avian superfamily Sylvioidea. Mol. Phylogenet. Evol. 38, 381–397.

Belfiore, N.M., Liu, L., Moritz, C., 2008. Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia: Geomyidae). Syst. Biol. 57, 294–310.

Beresford, P., Barker, F.K., Ryan, P.G., Crowe, T.M., 2005. African endemics span the tree of songbirds (Passeri): molecular systematics of several evolutionary 'enigmas'. Proc. R. Soc. Lond. B 272, 849–858.

Bonfield, J.K., Smith, K.F., Staden, R., 1995. A new DNA sequence assembly program. Nucleic Acids Res. 24, 4992–4999.

Cibois, A., Thibault, J.-C., Pasquet, E., 2007. Uniform phenotype conceals double colonization by reed-warblers of a remote Pacific archipelago. J. Biogeogr. 34 (7), 1155–1166.

Cramp, S., 1992. The Birds of the Western Palearctic, vol. 7. Oxford University Press, Oxford.

Del Hoyo, J., Elliott, A., Christie, D.A. (Eds.), 2006. Handbook of the Birds of the World. Old World Flycatchers to Old World Warblers, vol. 11. Lynx Edicions, Barcelona.

Devillers, P., Dowsett-Lemaire, F., 1978. African Reed Warblers (*Acrocephalus baeticatus*) in Kaouar (Niger). Le Gerfaut 68, 211–213.

Dickinson, E.C., 2003. The Howard & Moore Complete Checklist of the Birds of the World, third ed. Princeton University Press, Princeton, New Jersey.

Dickinson, E.C. 2007. Corrigenda 7. Available from: <http://birds-howardandmoore.org/content/howard-and-moore-corrigenda>.

Dowsett-Lemaire, F., Dowsett, R.J., 1987. European and African Reed Warblers, *Acrocephalus scirpaceus* and *A. baeticatus*: vocal and other evidence for a single species. Bull. Br. Ornithol. Club. 107 (2), 74–85.

Drummond, A.J., Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC 7, 214. Available from: <http://beast.bio.ed.ac.uk/Tracer>.

Edwards, S.V., Wilson, A.C., 1990. Phylogenetically informative length polymorphism and sequence variability in mitochondrial DNA of Australian songbirds (*Pomatostomus*). Genetics 126, 695–711.

Erard, C., Fry, C.H., Grimes, L.G., Irwin, M.P.S., Keith, S., Lack, P.C., Pearson, D.J., Tye, A., 1997. Sylviidae, Old World warblers. In: Urban, E.K., Fry, C.H., Keith, S. (Eds.), The Birds of Africa, vol. 5. Academic Press, San Diego, pp. 57–431.

Erixon, P., Svennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian Posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52 (5), 665–673.

Fleischer, R.C., Slikas, B., Beadell, J., Atkins, C., McIntosh, C.E., Conant, S., 2007. Genetic variability and taxonomic status of the Nihoa and Laysan Millerbirds. Condor 109, 954–962.

Fry, C.H., Williamson, K., Ferguson-Lees, I.J., 1974. A new subspecies of *Acrocephalus baeticatus* from Lake Chad and a taxonomic reappraisal of *Acrocephalus dumetorum*. Ibis 116, 340–346.

Grant, C.H.B., Mackworth-Praed, C.W., 1940. A new genus of African swamp warbler. Bull. Br. Ornithol. Club. 60, 91.

Haffer, J., 1991. Familie Sylviidae-Zweigsänger (Grasmücken und Verwandte). In: von Blotzheim, G.U.N., Bauer, K.M. (Eds.), Handbuch der Vögel Mitteleuropas, vol. 12. Aula-Verlag, Wiesbaden, pp. 11–21. and 208–216.

Hall, B.P., Moreau, R.E., 1970. An Atlas of Speciation in African Passerine Birds. Trustess of the British Museum (Natural History), London.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids. Symp. Ser. 41, 95–98.

Hartert, E., 1909. Die Vögel der paläarktischen Fauna. Verlag von R. Friedländer und Sohn, Berlin.

Helbig, A.J., Seibold, I., 1999. Molecular Phylogeny of Palearctic–African *Acrocephalus* and *Hippolais* Warblers (Aves: Sylviidae). Mol. Phylogenet. Evol. 11, 246–260.

Helbig, A.J., Kocum, A., Seibold, I., Braun, M.J., 2005. A multi–gene phylogeny of aquiline eagles (Aves: Accipitriformes) reveals extensive paraphyly at the genus level. Mol. Phylogenet. Evol. 35, 147–164.

Heslewood, M.M., Elphinstone, M.S., Tidemann, S.C., Baverstock, P.R., 1998. Myoglobin intron variation in the Gouldian Finch *Erythrura gouldiae* assessed by temperature gradient gel electrophoresis. Electrophoresis 19, 142–151.

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

Jobb, G., von Haeseler, A., Strimmer, K., 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4, 1471–2148.

Jobb, G. 2008. TREEFINDER version of October 2008. Munich, Germany. Distributed by the author at <www.treefinder.de>.

Johansson, U., Fjeldså, J., Bowie, R., 2008. Phylogenetic relationships within Passerida (Aves: Passeriformes): a review and a new molecular phylogeny based on three nuclear intron markers. Mol. Phylogenet. Evol. 48, 858–876.

Kennerley, P.R., Leader, P.J., 1992. The identification, status and distribution of small *Acrocephalus* warblers in eastern China. Hong Kong Bird Rep. 1991, 143–187.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86, 6196–6200.

Leisler, B., 1980. Morphological aspects of ecological specialization in bird genera. Oekol. Vögel. 2, 199–220.

Leisler, B., Heidrich, P., Schulze-Hagen, K., Wink, M., 1997. Taxonomy and phylogeny of reed warblers (genus *Acrocephalus*) based on mtDNA sequences and morphology. J. Ornithol. 138, 469–496.

Leisler, B., Ley, H.W., Winkler, H., 1989. Habitat, behaviour and morphology of *Acrocephalus* warblers: an integrated analysis. Ornis Scand. 20, 181–186.

Marova, I.M., Valchuk, O.P., Kvartalynov, P.V., Ivanitskij, V.V., 2005. On the taxonomic position and evolutionary interrelations of the Thick-billed Warbler, *Phragmaticola aedon* (based on ecological an ethological data). Alauda 73, 308.

Meise, W. 1976. In: Proceedings of the 16th International Ornithological Congress, vol. 1974, pp. 207–216.

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16, 1215.

Müller, K., 2005. SeqState – primer design and sequence statistics for phylogenetic DNA data sets. Appl. Bioinformatics 4, 65–69.

Müller, K., 2006. Incorporating information from length–mutational events into phylogenetic analyses. Mol. Phylogenet. Evol. 38, 667–676.

Notredame, C., Higgins, D., Heringa, J. 2000. T–Coffee: a novel method for multiple sequence alignments. J. Mol. Biol. 302, 205–217. Available from: <http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi>.

Nylander, J.A.A. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University. Available from: <http://www.abc.se/~nylander/>.

Ottosson, U., Bensch, S., Svensson, L., Waldenström, J., 2005. Differentiation and phylogeny of the olivaceous warbler *Hippolais pallida* species complex. J. Ornithol. 146, 127–136.

Parkin, D.T., Collinson, M., Helbig, A.J., Knox, A.G., Sangster, G., Svensson, L., 2004. Species limits in *Acrocephalus* and *Hippolais* warblers from the Western Palaearctic. Br. Birds 97, 276–299.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Round, P.D., Hansson, B., Pearson, D.J., Kennerley, P.R., Bensch, S., 2007. Lost and found: the enigmatic large-billed reed warbler *Acrocephalus orinus* rediscovered after 139 years. J. Avian Biol. 38, 133–138.

Schönwetter, M., 1979. Handbuch der Oologie. Band II. Akademie-Verlag, Berlin.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Sibley, C.G., Ahlquist, J.E., 1990. Phylogeny and Classification of Birds. A Study in Molecular Evolution. Yale University Press, New Haven and London.

Sibley, C.G., Monroe, B.L., 1990. Distribution and Taxonomy of Birds of the World. Yale Univ. Press, New Haven, CT.

Sibley, C.G., Monroe, B.L., 1993. A Supplement to Distribution and Taxonomy of Birds of the World. Yale Univ. Press, New Haven, CT.

Slade, R.W., Moritz, C., Heideman, A., Hale, P.T., 1993. Rapid assessment of single-copy nuclear DNA variation in diverse species. Mol. Ecol. 2, 359–373.

Sorenson, M.D., Quinn, T.W., 1998. Numts: a challenge for avian systematics and population biology. Auk 115, 214–221.

Svensson, L., Prys-Jones, R., Rasmussen, P.C., Olsson, U., 2008. Discovery of ten new specimens of large billed reed warbler *Acrocephalus orinus*, and new insights into its distributional range. J. Avian Biol. 39, 605–610.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Vaurie, C., 1959. The Birds of the Palearctic. Fauna Order Passeriformes. Witherby, London.

Voous, K.H., 1975. An aberrant Reed Warbler, or: on the inequality of genera in birds. Ardeola, 977–985.

Watson, G.E., Traylor, M.A., Mayr, E., 1986. Family Sylviidae, Old World Warblers. In: Mayr, E., Cottrell, G.W. (Eds.), Check-list of the Birds of the World, vol. 4. Mus. of Comp. Zool, Cambridge, MA., pp. 3–294.

Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D.G., Sites Jr., J.W., Townsend, T.M., Reeder, T.W., 2008. Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. Syst. Biol. 57, 420–431.

Williamson, K., 1968. Identification for ringers 1: the genera *Cettia*, *Locustella*, *Acrocephalus* and *Hippolais*, third ed. British Trust for Ornithology, Tring.

Wilkinson, R., Aidley, D.J., 1983. African Reed Warblers in northern Nigeria; morphometrics and the taxonomic debate. Bull. Br. Ornithol. Club. 103, 135–138.

Wolters, H.E., 1982. Die Vogelarten der Erde. 7. Lieferung, Parey, Hamburg/Berlin.

Molecular Phylogenetics and Evolution 58 (2011) 513–526

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Multilocus analysis of a taxonomically densely sampled dataset reveal extensive non-monophyly in the avian family Locustellidae

Per Alström [a,*], Silke Fregin [b], Janette A. Norman [c], Per G.P. Ericson [d], Les Christidis [c,e], Urban Olsson [f]

[a] Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, SE-750 07 Uppsala, Sweden
[b] Vogelwarte Hiddensee, Zoological Institute and Museum, Ernst Moritz Arndt University of Greifswald, 17489 Greifswald, Germany
[c] Department of Genetics, University of Melbourne, Parkville Vic. 3052 Australia and Sciences Department, Museum Victoria, GPO Box 666, Melbourne Vic. 3001, Australia
[d] Department of Vertebrate Zoology and Molecular Systematics Laboratory, Swedish Museum of Natural History, P.O. Box 50007, SE-104 05 Stockholm, Sweden
[e] National Marine Science Centre, Southern Cross University, NSW 2450, Australia
[f] Department of Zoology, University of Gothenburg, Box 463, SE-405 30 Göteborg, Sweden

ARTICLE INFO

Article history:
Received 12 October 2010
Revised 17 December 2010
Accepted 20 December 2010
Available online 15 January 2011

Keywords:
Species trees
Gene trees
Phylogeny
Concatenation
BEST
MrBayes
Classification

ABSTRACT

The phylogeny of most of the species in the avian passerine family Locustellidae is inferred using a Bayesian species tree approach (Bayesian Estimation of Species Trees, BEST), as well as a traditional Bayesian gene tree method (MrBayes), based on a dataset comprising one mitochondrial and four nuclear loci. The trees inferred by the different methods agree fairly well in topology, although in a few cases there are marked differences. Some of these discrepancies might be due to convergence problems for BEST (despite up to 1×10^9 iterations). The phylogeny strongly disagrees with the current taxonomy at the generic level, and we propose a revised classification that recognizes four instead of seven genera. These results emphasize the well known but still often neglected problem of basing classifications on non-cladistic evaluations of morphological characters. An analysis of an extended mitochondrial dataset with multiple individuals from most species, including many subspecies, suggest that several taxa presently treated as subspecies or as monotypic species as well as a few taxa recognized as separate species are in need of further taxonomic work.

1. Introduction

The avian family Sylviidae (“Old World warblers”) has long been recognized as one of the main passerine families, although the composition has varied among authors. Traditionally, a large number of taxa were included, e.g. 60 genera and 358 species in the classification of Watson et al. (1986). Sibley and Monroe (1990), based on the DNA–DNA hybridization work by Sibley and Ahlquist (1990), split off Cisticolidae from Sylviidae, and further divided Sylviidae into the subfamilies Megalurinae, Acrocephalinae and Sylviinae. This was followed by Dickinson (2003) and Bairlein et al. (2006). Later studies, based on DNA sequence data, revised this classification. Alström et al. (2006) and Johansson et al. (2008) proposed recognition of a number of well supported major clades at family level. These authors synonymized Sylviidae with the family Timaliidae (“babblers”). Gelang et al. (2009), again based on DNA sequence data, resurrected Sylviidae, but restricted it to a clade containing mainly traditional Timaliidae species.

The subfamily Megalurinae sensu Sibley and Monroe (1990) contained the genera *Megalurus*, *Cincloramphus*, *Eremiornis*, *Amphilais*, *Megalurulus*, *Buettikoferella*, *Chaetornis*, *Graminicola* and *Schoenicola*. In contrast, the family Megaluridae sensu Alström et al. (2006) and Johansson et al. (2008) comprised the genera *Megalurus*, *Bradypterus*, *Locustella* and *Dromaeocercus*, i.e. including three of the genera placed in Acrocephalinae by Sibley and Monroe (1990). Other DNA sequence studies have shown that *Cincloramphus* and *Schoenicola* form a clade with *Bradypterus* and *Megalurus* (Beresford et al., 2005), while *Graminicola* belongs to the babbler family Timaliidae (Alström et al., 2006; Gelang et al., 2009). Beresford et al. (2005) also revealed that the aberrant *Bradypterus victorini* is not related to Megaluridae/Megalurinae.

The name Locustellinae Bonaparte, 1854, has priority over Megalurinae Blyth, 1875 (Bock, 1994: p. 152), and thus the family name Locustellidae Bonaparte, 1854 is applied in the present paper for Megaluridae sensu Alström et al. (2006) and Johansson et al. (2008). The relationships within this family are poorly known. Drovetski et al. (2004) used mitochondrial ND2 to study the relationships of all *Locustella*, two Asian and three African *Bradypterus*, and two *Megalurus*. They found the Asian *Bradypterus* and *Megalurus pryeri* nested within *Locustella*, the African *Bradypterus* in a separate clade, and *M. gramineus* on a branch on its own.

* Corresponding author.
E-mail addresses: per.alstrom@slu.se (P. Alström), silke.fregin@uni-greifswald.de (S. Fregin), janorman@unimelb.edu.au (J.A. Norman), per.ericson@nrm.se (P.G.P. Ericson), les.christidis@gmail.com (L. Christidis), urban.olsson@zool.gu.se (U. Olsson).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.12.012

The species in Locustellidae are distributed across Africa, Eurasia and Australasia, frequenting mostly bushy, but sometimes also marshy, habitats from sea level up to above the tree limit (c. 4500 m in the Himalayas) (Bairlein et al., 2006). Most species are notoriously secretive and difficult to observe. All are non-descript, mostly various shades of brown above and at least slightly paler below; *Megalurus*, *Cincloramphus* and some *Locustella* are streaked above, some of these and some *Bradypterus* also on the underparts (Bairlein et al., 2006). *Cincloramphus cruralis* is exceptional in that the male is uniformly dark sooty brown below (Bairlein et al., 2006). Most species are fairly small, with an overall length of 13–16 cm, but some are considerably larger (22–28 cm in *Megalurus palustris*) (Bairlein et al., 2006). The songs are mostly simple but distinctive, and in general differ more than morphology among closely related species (Bairlein et al., 2006). Due to the generally cryptic plumages, there has been much confusion regarding species level taxonomy (e.g. Dickinson et al., 2000), and recent studies involving vocalizations and/or DNA have led to suggestions that some taxa currently treated as subspecies should be raised to the rank of species (e.g. Drovetski et al., 2004; Alström et al., 2008) as well as to the identification of a new cryptic species (Rasmussen et al., 2000).

In the present study, we infer the relationships of nearly all species in the family Locustellidae using one mitochondrial gene and four nuclear introns. We use traditional gene tree methods (Bayesian inference, maximum likelihood bootstrapping, parsimony bootstrapping) as well as a Bayesian species tree approach (Bayesian Estimation of Species Trees, BEST; Liu and Pearl, 2007; Liu, 2008) that accounts for lineage sorting processes that might produce discordance between gene trees. We also analyse mitochondrial DNA for a larger sample, comprising multiple individuals and several subspecies of polytypic species. A revised taxonomy is proposed based on our results.

2. Materials and methods

2.1. Study group

In total, we include 37 species from seven genera considered to belong to Locustellidae (=Megaluridae sensu Alström et al., 2006 and Johansson et al., 2008). Our sample comprises 16 species of *Bradypterus* plus cytochrome *b* (cyt*b*) sequences for three additional species (two from GenBank and one provided by Trevor Price and Udayan Borthakur; only two African and three Asian species are missing); all eight *Locustella* species; four *Megalurus* species plus cyt*b* for one more species (two species are lacking); both species of *Cincloramphus*; one of the two species of *Schoenicola*; and the monotypic genera *Dromaeocercus* and *Eremiornis*. For cyt*b*, we have in total 82 unique haplotypes, including 24 sequences from GenBank, comprising several taxa treated as subspecies of polytypic species. Sequences from four nuclear markers (ODC, myo, GAPDH, LDH) were obtained for most taxa (see Appendix A for details regarding loci coverage across the taxa).

Species level taxonomy follows Dickinson (2003) and Bairlein et al. (2006), with the exception of the recognition of *Bradypterus thoracicus kashmirensis* as a distinct species, based on a study of morphology, vocalizations and mitochondrial DNA (Alström et al., 2008).

2.2. DNA extraction and sequencing

DNA was extracted from blood, feathers, or muscle, using QIA Quick DNEasy Kit (Qiagen, Inc.) according to the manufacturer's instruction, but with 30 µl 0.1% DTT added to the initial incubation step of the extraction of feathers. We sequenced five loci: the main part of the mitochondrial cytochrome *b* gene and part of the flanking tRNA-Thr (cyt*b*); the nuclear ornithine decarboxylase exon 6 (partial), intron 6, exon 7, intron 7 and exon 8 (partial) (ODC); the entire nuclear myoglobin intron 2 (myo), the nuclear glyceraldehyde-3-phosphodehydrogenase intron 11 (GAPDH), and the complete nuclear lactate dehydrogenase intron 3 (LDH). Amplification and sequencing of cyt*b* and myo followed the protocols described in Olsson et al. (2005), of ODC Allen and Omland (2003), of GAPDH Fjeldså et al. (2003), and of LDH Fregin et al. (2009). Cyt*b* was amplified as one fragment to decrease the risk of amplifying nuclear pseudocopies (e.g. Sorensen and Quinn, 1998). DNA was also extracted from one museum specimen (*Schoenicola brevirostris*). For extraction, PCR-amplification, and sequencing procedures from this one, the procedures described in Irestedt et al. (2006) were followed, with specially designed primers obtainable from the authors upon request. All new sequences have been deposited in GenBank (Appendix A).

2.3. Phylogenetic analyses

Sequences were aligned using MegAlign 4.03 in the DNASTAR package (DNAstar Inc.); some manual adjustment was necessary for the non-coding sequences. For the nuclear loci, haplotypes were not separated, but coded as ambiguous bases.

Gene trees were estimated by Bayesian inference (BI) using MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001, 2005) according to the following: (1) All loci were analysed separately (single-locus analyses, SLAs). (2) Sequences were also concatenated, either all nuclear loci, or all loci together. In the multilocus analyses, the data were either (a) partitioned by locus, using rate multipliers to allow different rates for the different partitions (Nylander et al., 2004; Ronquist and Huelsenbeck, 2003), or (b) unpartitioned, using a homogeneous model for the entire dataset. In the analyses of all loci, species with missing data were included or excluded in various constellations. Ambiguous base pairs and indels were treated as missing data, but indels were plotted on the trees a posteriori. As outgroups, two species belonging to the family Bernieridae (*Hartertula flavoviridis* and *Thamnornis chloropetoides*) were chosen, as this family has been suggested to be sister to Locustellidae (Beresford et al., 2005; Johansson et al., 2008). Analyses were also run with 28 outgroup species, representing all families in the superfamily Sylvioidea (Alström et al., 2006; Johansson et al., 2008).

Appropriate substitution models were determined based on the Akaike Information Criterion (Akaike, 1974) and a hierarchical likelihood ratio test (Posada and Crandall, 1998), both calculated using MrModeltest2 (Nylander, 2004) in conjunction with PAUP* (Swofford, 2002). For all loci, posterior probabilities (PPs) were calculated under the general time-reversible (GTR) model (Lanave et al., 1984; Tavaré, 1986; Rodríguez et al., 1990), assuming rate variation across sites according to a discrete gamma distribution with four rate categories (Γ; Yang, 1994) and, for the cyt*b* data, also an estimated proportion of invariant sites (I; Gu et al., 1995). Default priors in MrBayes were used. Four Metropolis-coupled MCMC chains with incremental heating temperature 0.1 or 0.2 were run for 10–30 $\times$ 10^6 generations and sampled every 1000 generations. Chain likelihood and other parameter values and effective sample sizes (>200, generally >1000) were inspected in Tracer 1.5.0 (Rambaut and Drummond, 2009). The first 25% of the generations were discarded as "burn-in", well after stationarity of chain likelihood values had been established, and the posterior probability was estimated for the remaining generations. Every analysis was run at least twice, and the topologies and posterior probabilities compared by eye.

Species tree analysis was performed using Bayesian Estimation of Species Trees (BEST) 2.3 (Liu and Pearl, 2007; Liu, 2008). Only

species with complete data were included. Two long analyses were run, each with four Metropolis-coupled MCMC chains running 1×10^9 cycles. In addition, eight shorter analyses, each c. 7×10^7–1×10^8 replicates, were run. All analyses were sampled every 1000 generations, and the incremental heating temperature was set to 0.1. The theta prior was set to invgamma (3, 0.003) and the GeneMu prior to uniform (0.5, 1.5). The posterior distribution was summarized based on the generations with the highest, seemingly stable, likelihood values. *H. flavoviridis* and *T. chloropetoides* were again used as outgroups.

In addition, clades B, C and D identified in the BI (Figs 1–3) were analysed separately by BEST in order to try and get better convergence than in the more extensive BEST analyses. These analyses were run for 2×10^8 generations, all else being equal to the other BEST. The outgroups were the same as in the other BEST analyses.

Maximum likelihood bootstrapping (MLB) (1000 replicates) was performed on the complete dataset in Treefinder (version of October 2008; Jobb et al., 2004; Jobb, 2008) using default settings and a uniform GTR+Γ+I model. Parsimony bootstrapping (MPB) was performed in PAUP* (Swofford, 2002) on the complete dataset:

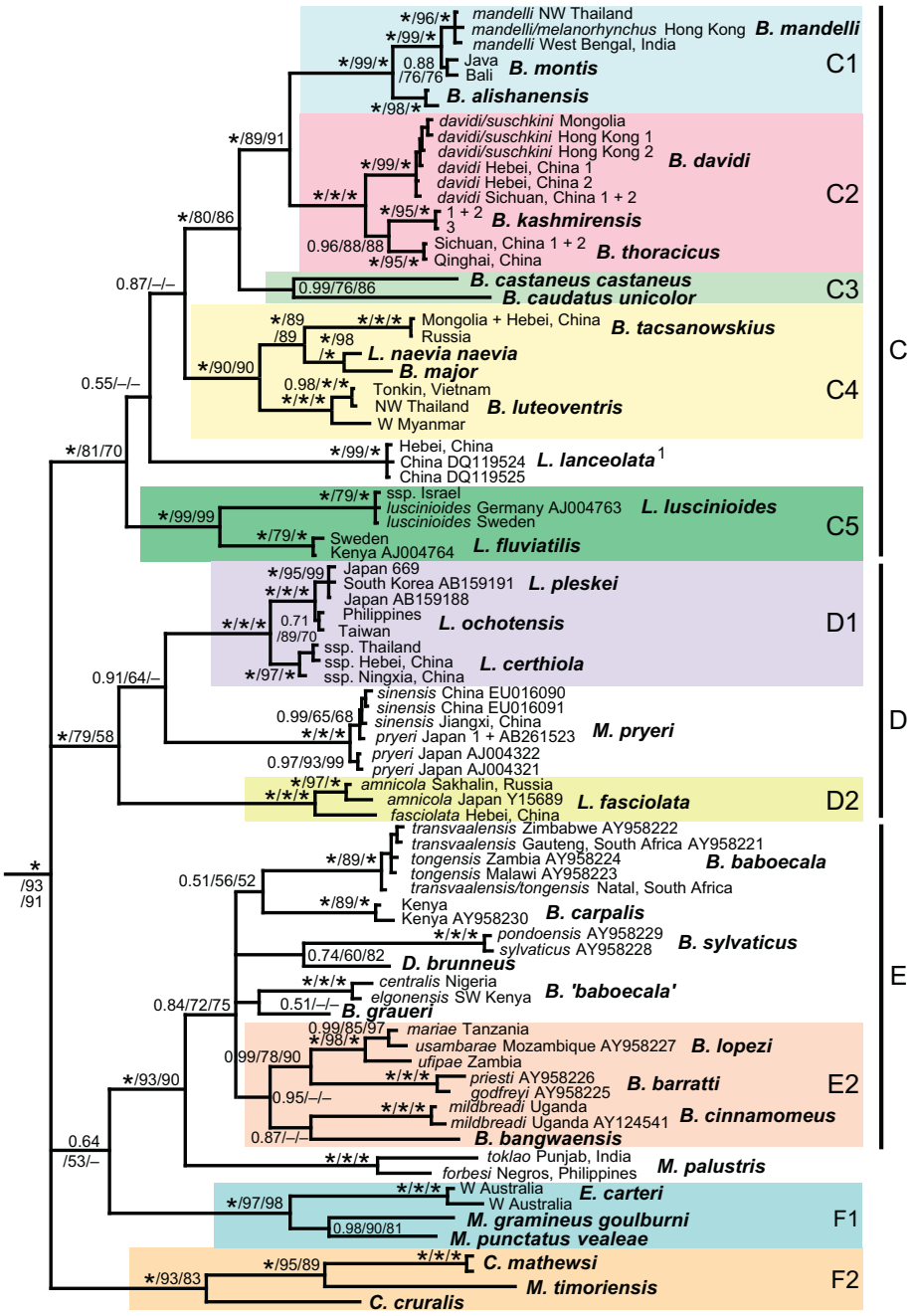

Fig. 1. Majority rule (50%) consensus tree of Locustellidae based on unique mitochondrial cytochrome *b* haplotypes, inferred by Bayesian inference under the GTR+Γ+I model. Posterior probabilities, and maximum likelihood and parsimony bootstrap values are indicated at the nodes, in this order; an asterisk represents posterior probability 1.0 or bootstrap 100%. The species for which no subspecific names are given are monotypic (except *L. lanceolata*, see below). Bars and colour shading delimit clades discussed in text. *B.* = *Bradypterus*, *C.* = *Cincloramphus*, *D.* = *Dromaeocercus*, *E.* = *Eremiornis*, *L.* = *Locustella*, and *M.* = *Megalurus*. Numbers after names are sample identifiers (e.g. *davidi* Sichuan, China 1 + 2 means *davidi* samples 1 and 2 from Sichuan, China [same haplotype], and *sinensis* China EU016090 refers to GenBank number of sequence previously used in another study; see Appendix A). [1]On geographical grounds, most likely nominate subspecies, but samples collected during migration, so subspecies *hendersonii* cannot be eliminated.

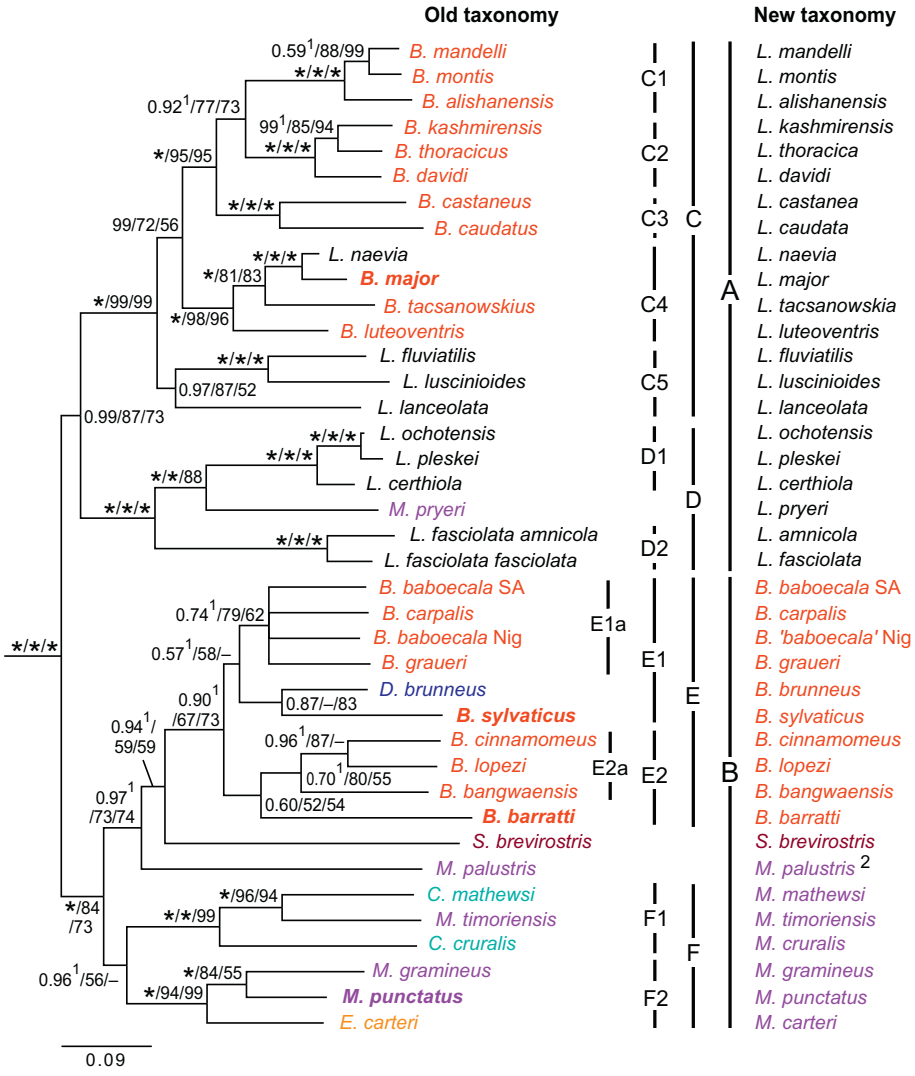

Fig. 2. Majority rule (50%) consensus tree of Locustellidae based on concatenated nuclear ODC, myoglobin, LDH and GAPDH and mitochondrial cytochrome *b* sequences, inferred by Bayesian inference, analysed in five partitions (four nuclear loci GTR+Γ, cytochrome *b* GTR+Γ+I). Colours of names indicate genus according to old taxonomy (Dickinson 2003; left) and new taxonomy proposed here (right; see Fig. 1 for explanation of abbreviations of generic names; *S.* = *Schoenicola*). Labelled bars denote clades as in Fig. 1. The four species for which only cytochrome *b* is available are in bold type. Posterior probabilities, and maximum likelihood and parsimony bootstrap values are indicated at the nodes, in this order; an asterisk represents posterior probability 1.0 or bootstrap 100%. *B. baboecala* SA and *B. baboecala* Nig refer to samples from South Africa (*transvaalensis/tongensis*) and Nigeria (*centralis*), respectively. [1]Node affected differently by different types of analyses (see Table 1). [2]See Section 4 for recognition of non-monophyletic *Megalurus*.

heuristic search strategy, 1000 replicates, starting trees obtained by stepwise addition (random addition sequence, 10 replicates), TBR branch swapping, MulTrees option not in effect (only one tree saved per replicate).

Bayes factors (Newton and Raftery, 1994; Kass and Raftery, 1995) were calculated in Tracer 1.5.0 (Rambaut and Drummond, 2009) for comparisons of alternative hypotheses in some BI analyses.

GTR+Γ+I corrected pairwise divergences for the cyt*b* dataset (excluding outgroup species) were calculated in Treefinder (version of October 2008; Jobb et al., 2004; Jobb, 2008). Positions where one or more taxa had ambiguous nucleotides were deleted from the matrix, and incomplete sequences were excluded, or the ends were trimmed, so that all sequences used in the comparisons comprised 982 base pairs.

2.4. Summary of abbreviations

BI – Bayesian inference (MrBayes); BIC – Bayesian inference (MrBayes) of concatenated sequences; cyt*b* – cytochrome *b* gene; GAPDH – glyceraldehyde-3-phosphodehydrogenase intron 11; LDH – lactate dehydrogenase intron 3; MLB – maximum likelihood bootstrap; MPB – parsimony bootstrap; myo – myoglobin intron 2; ODC – ornithine decarboxylase (mainly) introns 6–7; PP – posterior probability; SLA – single-locus analysis.

3. Results

3.1. Sequence characteristics

We obtained a contiguous ≤730 base pair (bp) stretch of the ODC, ≤709 bp of myo, ≤510 bp of the LDH, ≤375 bp of the GAPDH, and ≤1038 bp of cyt*b*. No unexpected stop codons, indels, or distinct double peaks in the chromatograms that would indicate the presence of nuclear pseudogenes were found in the coding cyt*b* sequences, except for one sequence from GenBank: the *Locustella fasciolata amnicola* sequence with Genbank No. Y15689 contains a stop codon (AGA). However, this individual proved to be sister to the other individual of the same taxon, and these two form the sister clade to *L. fasciolata fasciolata*, so either the paralogue is of recent origin or the G in the stop codon is a misreading in

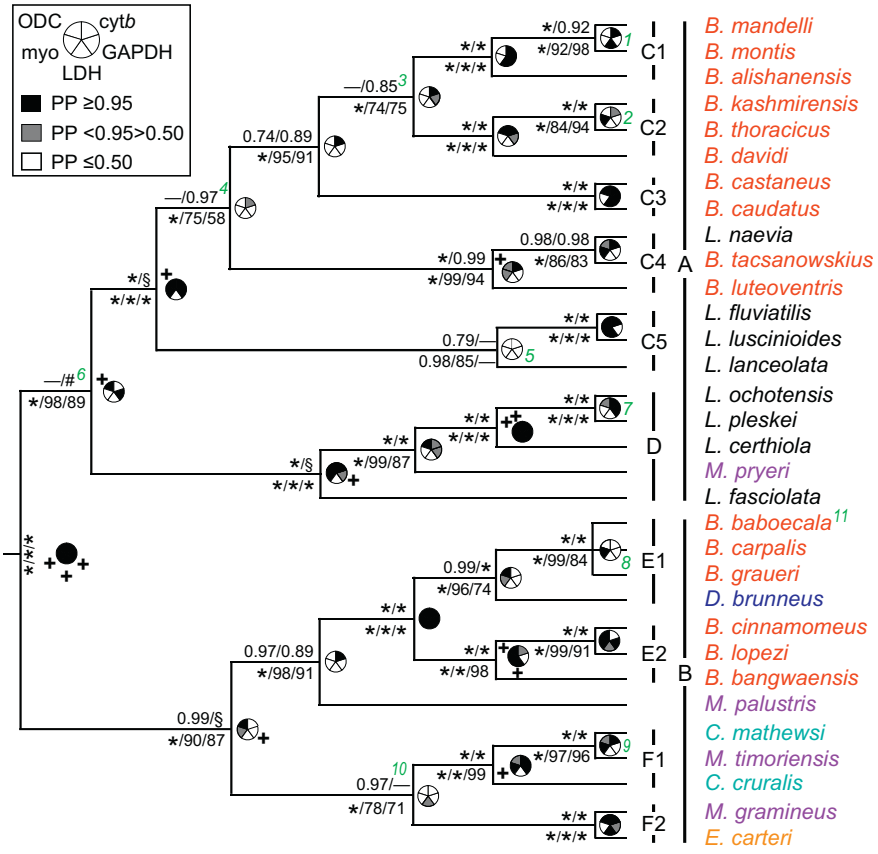

Fig. 3. Consensus phylogeny of Locustellidae based on analyses of nuclear ODC, myoglobin, LDH and GAPDH and mitochondrial cytochrome *b* sequences. Only species for which all loci are available are included. Values above branches represent species tree inferences (BEST posterior probability: entire dataset/separate analyses of clades B–D), and values below branches represent analyses based on concatenation (Bayesian posterior probability [unpartitioned: GTR+Γ+I]/maximum likelihood bootstrap/parsimony bootstrap). * indicates posterior probability 1.0 or bootstrap 100%; § indicates result from BEST not applicable, since only two distant outgroups were used in separate analyses of clades B–D; # indicates that no BEST was performed on clade A separately. Pie charts at nodes denote support in single-locus analyses (see explanation in upper left corner); + indicates further support by indel (in the locus whose pie is adjacent to the +; cf. Supplementary Fig. 3). Labelled bars denote clades as in Figs. 1 and 2. Colours of names indicate genus according to old taxonomy (Dickinson, 2003); see Figs. 1 and 2 for explanation of abbreviations of generic names. [1]In single-locus analysis of both cyt*b* and LDH this clade has PP 1.0, whereas in analysis of ODC clade (*B. mandelli*, *B. alishanensis*) has PP 0.99. [2]In single-locus analysis of cyt*b* and myo, this clade has PP 0.86 and 0.99, respectively, whereas in analysis of ODC, clade (*B. davidi*, *B. kashmirensis*) has PP 1.0. [3]In BEST analysis of entire dataset, clade C2 is sister to clade C3 with PP 0.80. [4]In BEST analysis of entire dataset, clades C1–C3, C4 and C5 form a trichotomy. [5]Not supported by any single-locus analysis. [6]In BEST of complete dataset, clade D is sister to the rest with PP 0.67. [7]In myo tree, *L. ochotensis* and *L. certhiola* are sisters with PP 1.0. [8]In LDH tree, clade (*B. baboecala*, *B. graueri*) has PP 0.97, while in myo tree (*B. baboecala*, *B. carpalis*) has PP 0.94. [9]In GAPDH tree, *C. cruralis* and *M. timoriensis* are sisters with PP 0.99. [10]See text. [11]*B. baboecala* concerns subspecies *transvaalensis/tongensis* from Natal, South Africa.

the original sequence. Including all outgroup taxa, the aligned ODC sequences comprise 762 characters, of which 203 (27%) are parsimony informative; myo 758 characters, 151 (20%) parsimony informative; LDH 541 characters, 164 (30%) parsimony informative; GAPDH 400 characters, 95 (24%) parsimony informative; and cyt*b* 1078 characters, 445 (41%) parsimony informative. Including all outgroups, the combined ODC, myo, LDH and GAPDH (hereafter nuclear) data set contains 2461 characters, of which 613 (25%) are parsimony informative, and the total data set comprises 3539 characters, of which 1058 (30%) are parsimony informative. Excluding the outgroup taxa, the aligned ODC sequences comprise 734 characters, of which 73 (9.9%) are parsimony informative; myo 709 characters, 43 (6.1%) parsimony informative; LDH 513 characters, 40 (7.8%) parsimony informative; GAPDH 379 characters, 35 (9.2%) parsimony informative; and cyt*b* 1038 characters, 348 (33.5%) parsimony informative. Excluding outgroups, the combined ODC, myo, LDH and GAPDH (hereafter nuclear) dataset contains 2335 characters, of which 191 (8.2%) are parsimony informative, and the total dataset comprises 3373 characters, of which 539 (16.0%) are parsimony informative. The cyt*b* dataset comprising multiple samples of most species includes 384 parsimony-informative characters (37.0%).

3.2. Single-locus analyses

The tree containing multiple cyt*b* haplotypes for most Locustellidae species, including several subspecies and species for which only cyt*b* is available, is overall well resolved (88% of species nodes), except for some, mostly deep, internal nodes (Fig. 1).

The trees based on single-locus analyses (hereafter SLAs) of single sequences per species vary in resolution: 85% of the nodes are bifurcating in the cyt*b* tree, 75% in the ODC tree, 59% in the myo tree, 68% in the LDH tree, and 55% in the GAPDH tree (Supplementary Fig. 1; see also Fig. 3, where SLAs are shown in pie charts). Although the resolution varies among these trees, they generally agree fairly well, and there are few strongly supported topological conflicts. Only four conflicting reconstructions receive ≥0.95 posterior probability (PP) in different trees: (1) *Bradypterus thoracicus* and *Bradypterus kashmirensis* are sisters according to myo (0.99) and cyt*b* (0.86; 0.96 in analysis with multiple individuals, Fig. 1), while *B. kashmirensis* and *Bradypterus davidi* are sisters according to ODC (1.0); (2) *Bradypterus mandelli* and *Bradypterus montis* are sisters according to LDH and cyt*b* (both 1.0), whereas *B. mandelli* and *Bradypterus alishanensis* are sisters according to ODC (0.99); (3) *Locustella ochotensis* and *Locustella pleskei* are sisters in the

ODC (0.57), LDH (0.98) and cyt*b* trees (1.0), while *L. ochotensis* and *Locustella certhiola* are sisters in the myo tree (1.0); and (4) *Cincloramphus mathewsi* and *Megalurus timoriensis* are sisters according to ODC (0.87), myo (1.0) and cyt*b* (1.0), whereas *C. cruralis* and *M. timoriensis* are sisters according to GAPDH (0.99).

3.3. Concatenated multilocus analyses

In the BI analyses with 28 outgroup species, representing all major clades in Sylvioidea, the ingroup is shown to be monophyletic, with strong support (Supplementary Fig. 3). Within Locustellidae, the BI trees based on concatenation of all loci vary among analyses in topology and support for certain clades, depending on partitioning of data, and inclusion or exclusion of species with missing data (Table 1). For example, the BIC including all species (also those with incomplete data) in five partitions results in a tree with all except two nodes bifurcating, and all of these except nine having PP $\geq$0.95 (Fig. 2), whereas the unpartitioned BIC including only species with complete data results in a tree with all except one node bifurcating with PP $\geq$0.95 (Fig. 3). Bayes factor comparisons of the partitioned and unpartitioned analyses (all else being equal) are shown in Table 1. Also the MLB and MPB trees vary depending on whether species with missing data are included or excluded (Figs. 2 and 3).

Of the 29 clades recovered in the unpartitioned BIC of species with complete data (Fig. 3), five are found in only one of the SLAs, six in two SLAs, seven in three SLAs, eight in four SLAs, and two in all five SLAs (cf. Fig. 3 and Supplementary Fig. 1). The clade comprising C1–C4, which is recovered in only one of the SLAs, has considerably higher posterior probability in the analysis of all loci than in the SLA (1.0 and 0.82, respectively). The inclusion of *Locustella lanceolata* in clade 5 receives PP >0.95, despite that it is not found in this clade in any of the SLAs, and is poorly supported in this clade by MLB and MPB.

3.4. Species tree analyses

The BEST analyses of the complete dataset (Fig. 3) had convergence problems, despite a large number (up to 1×10^9) of iterations. The analysis with the highest likelihood values reached a plateau after c. 7×10^8 generations, but then dipped again after c. 9×10^8 generations, so it is uncertain if it ever reached its target distribution. The other analysis of the complete data was still rising in likelihood values at the end of the analysis. Some of the eight shorter analyses (c. 7×10^7–1×10^8 replicates) appeared to reach stationarity, based on their likelihood plots, but all had significantly lower likelihoods than the best 1×10^9 run; the latter is strongly favoured over the shorter run with the highest likelihood (ln Bayes factor: 115; cf. Kass and Raftery, 1995). BEST analyses of subsets of species, corresponding to clades B, C and D in the BIC (Fig. 3), appeared to converge considerably faster. In general, the results of these analyses are similar to those of the BEST of all species. However, there are some pronounced differences between the two types of BEST analyses (most notably at the nodes indicated by 3, 4, 5 and 10 in Fig. 3).

Notwithstanding the discrepancies between the different BEST analyses, the BEST phylogeny conforms rather well in topology with the BIC trees, although the support values are lower on average in the BEST than in the BIC (Fig. 3). However, in a few cases either of the two types of BEST analyses fails to recover a clade found with strong support in the BIC (indicated by 3, 4, 5, 6 and 10 in Fig. 3). One of these clades (5) is not recovered in any SLA, two (4, 10) are found with PP <0.95 in single SLAs, and two (3, 6) are recovered in two SLAs (one or both with PP $\geq$0.95). All of these cases concern nodes in which the two types of BEST analyses disagree (except node 6, which was only analysed in the complete dataset).

3.5. Major clades in Locustellidae

The following refers to the tree based on the BIC of all loci and all species (Fig. 2) and the species tree based on fewer species (Fig. 3). In the BIC tree, Locustellidae is separated into two major, well supported clades (A and B), whereas the BEST analysis of the complete dataset does not recover clade A (no separate BEST analysis of clade A was done). Clade A comprises all of the *Locustella* (Eurasia: Palearctic), Asian (Oriental) *Bradypterus* and *Megalurus pryeri* (Asia: Palearctic), while clade B contains the African *Bradypterus*, the monotypic Malagasy genus *Dromaeocercus*, the African *Schoenicola brevirostris*, the two Australian species of *Cincloramphus*, the monotypic Australian genus *Eremiornis* and four species of *Megalurus* (south Asia to Australasia). Clade A is further divided into clades C and D, which are both strongly supported in all analyses. The former includes a mix of *Locustella* and *Bradypterus*

Table 1
Comparison of Bayesian inference (MrBayes) of all five loci concatenated. "All species" refers to analyses including also taxa for which one or more loci is missing (*L. fasciolata amnicola*, *B. sylvaticus*, *B. barratti*, *S. brevirostris* and *M. punctatus*) and *B. baboecala* Nig (=Nigeria). "Species with complete data" refers to analyses excluding these taxa. "Partitioned" refers to analyses in five partitions (four nuclear loci GTR+Γ, cytochrome *b* GTR+Γ+I); "unpartitioned" refers to analyses of unpartitioned data (GTR+Γ+I). Bayes factors for comparisons are given at bottom of table.

Clade	All species		Species with complete data	
	Partitioned	Unpartitioned	Partitioned	Unpartitioned
(*B. mandelli*, *B. montis*)	0.59	1.0	0.57	1.0
(*B. thoracicus*, *B. kashmirensis*)	0.99	1.0	0.99	1.0
(C1, C2)	0.92	1.0	0.91	1.0
E	0.90	0.97	1.0[b]	1.0[b]
E1	0.57	0.67	1.0[c]	1.0[c]
E1a	0.74	0.81	1.0[a]	1.0[a]
E2a	0.70	0.85	1.0	1.0
(*B. cinnamomeus*, *B. lopezi*)	0.96	0.98	1.0	1.0
(*S. brevirostris*, E)	0.94	$\leq$0.50	n.a.[d]	n.a.[d]
(*M. palustris*, (*S. brevirostris*, E))	0.97	0.70	1.0[d]	1.0[d]
F	0.96	0.80	1.0[e]	1.0[e]
	ln Bayes factors = 451 in favour of partitioned		ln Bayes factors = 211 in favour of unpartitioned	

[a] *B. baboecala* Nig (=Nigeria, subspecies *centralis*) not included.
[b] *B. baboecala* Nig, *B. sylvaticus* and *B. barratti* not included.
[c] *B. baboecala* Nig and *B. sylvaticus* not included.
[d] *S. brevirostris* not included.
[e] *M. punctatus* not included.

nested among each other, while the latter contains five *Locustella* and, nested among them, *M. pryeri*.

Clade B is divided into clades E and F, with *S. brevirostris* and *M. palustris* as sequential sisters to clade E. The positions of *S. brevirostris* and *M. palustris* vary among analyses, and due to missing data, the former was excluded from the species tree analyses. Clade E is well supported and comprises the African *Bradypterus* and, nested among them, *Dromaeocercus brunneus*. The support for clade F varies among analyses (none to strong) and is only recovered in one of the SLAs.

3.6. Minor clades in Locustellidae

The following refers to the BIC tree of all loci and all species (Fig. 2) and the species tree based on fewer species (Fig. 3), unless otherwise noted. All of clades C1–C5 are well supported, except for the inclusion of *L. lanceolata* in clade C5 (see above). All of them, except C4, contain either *Locustella* or Asian *Bradypterus* species, while C4 includes both genera. As has already been noted, the topologies of clades C1 and C2 are incongruent between different SLAs. For *Bradypterus major* only cyt*b* is available (from a museum specimen from 1931; Appendix A), and accordingly, its position in clade C4 rests on this locus alone.

The relationships within clade D are strongly supported, although the sister relationship between *L. ochotensis* and *L. pleskei* is contradicted by the myo analysis. Clades E1 and E2, which comprise African *Bradypterus* and the Malagasy *Dromaeocercus*, are both well supported in the analyses of the species for which all loci are available. However, inclusion of two species for which only cyt*b* sequences are available (*Bradypterus sylvaticus* and *Bradypterus barratti*) markedly reduces the support for these clades (Fig. 2).

In a BIC analysis of cyt*b* including one sequence per species, *M. palustris* is sister to *Bradypterus graueri* in clade E with poor support (PP 0.66; not shown) and with a branch 2.5 times as long as any other branch in clade E. In contrast, inclusion of two different subspecies of *M. palustris* place these two in a sister clade to clade E (Supplementary Fig. 1a), as is also the case in the cyt*b* tree containing multiple haplotypes (Fig. 1).

Clades F1 and F2 are both strongly supported in all analyses, also when *Megalurus punctatus*, for which only cyt*b* is available, is included (Figs 2 and 3). The former clade comprises one of the two *Cincloramphus* and one *Megalurus* as sisters and the other *Cincloramphus* as sister to these, while the second clade contains two *Megalurus* as sisters and the monotypic genus *Eremiornis* as sister to these.

3.7. Indels

Several clades are supported by apparently synapomorphic indels in the alignments of the nuclear loci (Fig. 3, Supplementary Figs. 2 and 3). Within Locustellidae, two clades have two indels each, and another eight clades have one indel each. All of the clades supported by indels have PP 1.0 in the BIC of all loci.

3.8. Intra- and interspecific cytochrome b divergences

The variation in cyt*b*, as indicated by branch lengths (Fig. 1) and GTR+Γ+I corrected distances, is generally low within the species for which we have multiple samples, including the polytypic species (0.0–1.2%, mean 0.5% ± 0.4; $n = 33$ pairwise comparisons, excluding the cases below) compared to the taxa treated as different species (4.0–29.4%, mean 19.3% ± 4.4; $n = 1435$ pairwise comparisons, excluding the cases below). However, there are a few cases of unexpectedly large intraspecific divergences (Table 2). In contrast, two pairs of taxa that are usually treated as separate species differ comparatively slightly (Table 2).

Table 2
Large intraspecific or small interspecific cytochrome *b* divergences (GTR+Γ+I corrected, 982 base pairs). Mean and standard deviation in parentheses; numbers given in parentheses in left column.

Taxa	GTR+Γ+I corrected distances (%)
Intraspecific divergences	
B. baboecala tongensis/transvaalensis (South Africa; 1) vs. *B. b. centralis* (Nigeria; 1) and *B. b. elgonensis* (Kenya; 1)[a]	10.2 – 10.9 (10.6 ± 0.5)
Locustella fasciolata fasciolata (1) vs. *L. f. amnicola* (1[b])	5.0
B. lopezi mariae (Tanzania; 1) vs. *B. l. ufipae* (Zambia; 1)	3.6
M. palustris toklao (India; 1) vs. *M. p. forbesi* (Philippines; 1)[c]	6.5
B. luteoventris[d] W Myanmar (1) vs. Thailand (1) and Vietnam (1)	2.7 (±0)
E. carteri[d] from same locality (<2.5 km apart; 2)	2.0[e]
Interspecific divergences	
B. mandelli[d] (2) vs. *B. montis*[d] (2)	1.3–1.7 (1.5 ± 0.2)
L. pleskei[d] (3) vs. *L. ochotensis*[d] (2)	0.5–0.8 (0.7 ± 0.2)

[a] Non-sisters.
[b] Sequence from GenBank (Y15689) suspected of being a nuclear copy not included.
[c] Based on 898 base pairs.
[d] Monotypic.
[e] Based on 714 base pairs.

4. Discussion

4.1. Comparison of methods

In the present study there is comparatively little incongruence between different SLAs, with only four nodes in the Locustellidae clade being strongly incongruent. Accordingly, as expected, there is little difference between the trees reconstructed via species tree approaches and concatenation, and no signs of the former receiving additional signal from the likelihood function of gene trees given a species tree (cf. Edwards et al., 2007; Brumfield et al., 2008; Liu et al., 2008; Liu and Edwards, 2009; Edwards, 2009). Although several nodes have low statistical support in the BEST trees, with few exceptions they nevertheless recover the same topology as the trees inferred via concatenation. In spite of the slight differences between the species tree and gene tree approaches, we consider the former to be a step forward compared to the latter, since it accounts for the ubiquitous heterogeneity in gene trees, thereby providing more realistic support than concatenation for nodes with incongruence among loci or instances where all or most of the signal in a multilocus analysis comes from a single locus.

In the trees inferred from the BEST of the complete dataset, 92% of the nodes have PP ≥0.95 (mean PP 0.98; nodes with ≤0.5/50% support excluded), whereas BIC recover all nodes with PP >0.95 (mean 1.0). Accordingly, our results confirm the prediction that statistical confidence is generally lower in species trees than in trees estimated via concatenation (Edwards, 2009), as has also been found in other empirical studies employing BEST (e.g. Belfiore et al., 2008; Thomson et al., 2008). Although we found BEST to yield PP ≥0.95 for nodes that were only strongly supported in one SLA, at least three independent SLAs with PP ≥0.50 or two with PP ≥0.95 for a certain node were required for BEST to consistently attain PP ≥0.95 for that node (even if strongly contradicted by one other SLA). Similar results were obtained by Edwards et al. (2007) based on simulated data. However, Brumfield et al. (2008) and Liu et al. (2008) reported BEST inferring a species tree that was corroborated by independent data despite that this was not found by any of five SLAs. In spite of the increase in phylogenetic signal in species tree analyses compared to concatenation (e.g. Edwards et al., 2007; Brumfield et al., 2008; Eckert and Carstens, 2008; Liu and

Edwards, 2009; Edwards, 2009), it seems advisable to treat clades that are found in only one (or even no) SLA with caution, even when these have high BEST support. In the present study, this concerns nodes 4 and 10 in Fig. 3, although in these cases the BEST support varies between the complete and incomplete datasets (see below).

A few species tree reconstructions need to be commented on. The nodes marked by 3 and 4 in Fig. 3 are not recovered by the BEST of the complete dataset, whereas both these clades are inferred in one or two SLAs (and not contradicted by any other SLA), in all analyses of concatenated data, and in the BEST of clade C on its own. Moreover, the most basal node in clade A (indicated by 6 in Fig. 3), which is well supported by two SLAs, one indel, and all analyses of concatenated data is not recovered by BEST (only complete dataset analysed). In contrast, BEST places clade D as sister to the rest of the ingroup, with low support. With respect to clade A, the topology inferred by the majority of the analyses seems more probable based on morphology and biogeography (cf. Bairlein et al., 2006). In all these cases, it seems possible (even likely) that the BEST analyses of the complete data have not reached their target distributions (see Section 4.2, below).

The BIC analyses result in varying support depending on whether the data are partitioned or not. It is not evident which analysis is better. With respect to the analyses of all species, Bayes factor comparisons strongly favour the partitioned analysis over the unpartitioned one, whereas the opposite is true in the analyses including only species with complete data (all else being equal). It could be argued, however, that partitioned analyses are generally superior to unpartitioned analyses (e.g. Brown and Lemmon, 2007; McGuire et al., 2007), especially in cases where different loci have markedly different phylogentic signal; in the present study, cyt*b* is much more informative than the nuclear loci.

In the BIC analyses, inclusion of species for which only cyt*b* is available has varying effects in different parts of the tree. Inclusion of two such species (*B. sylvaticus*, *B. barratti*) negatively affects the support for several nodes in clade E, whereas inclusion of two other species with only cyt*b* data that are inferred to belong in clades C4 (*B. major*) and F2 (*M. punctatus*), respectively, do not appear to reduce the support for any neighbouring nodes (cf. Figs 2 and 3). These differences might be the result of the different lengths of these sequences: the sequences for *B. sylvaticus* and *B. barratti* are only 603 base pairs, whereas for *B. major* and *M. punctatus* they are 711 and 716 base pairs, respectively.

4.2. Convergence in BEST analyses

The BEST analyses of the complete dataset obviously had convergence problems, despite the large number of iterations. Even the longest run with the highest likelihood was fluctuating markedly near the end, while the other 1×10^9 run was still climbing when it terminated. The shorter BEST results from the complete dataset all had lower likelihood values, and therefore appeared not to have reached their target distributions, despite some having apparently spuriously stationary likelihood values. The differences in topology and support between the BEST of the complete dataset and the separate analyses of clades B, C and D might be due to convergence problems, especially in the more extensive dataset.

Convergence problems for BEST have been reported in other empirical studies. In a BEST of 162 genes from five species of *Zea* maize, Cranston et al. (2009) failed to reach convergence in 1.6×10^9 iterations, and Linnen and Farrell (2008) reported lack of convergence in multiple 50×10^6 generation runs for a *Neodiprion* saw fly dataset. These and the present results suggest that BEST might need to be run exceedingly long to reach the proper target distribution. Our results also emphasize that it is advisable to do multiple analyses of the same dataset to ascertain that convergence has been reached. If other analyses suggest the presence of some well corroborated monophyletic subgroups, analysing these separately, as also tested here, is likely to help BEST converge more quickly. Cranston et al. (2009) suggest that it might be possible to increase the rate of convergence by exploring MCMC parameters, using different proposal mechanisms, or perhaps by inferring starting parameters for the individual genes before beginning the joint analysis. An alternative solution might be to vary the population size (θ) prior. This has proven helpful in a study of *Sceloporus* fence lizards, in which only analyses with higher θ values ($\geq$0.015) converged (Leaché, 2009).

4.3. Phylogeny of Locustellidae

The phylogenetic hypothesis in Fig. 3 is mostly well supported by the data, although resolution of some internal nodes is uncertain. Clade A, which contains the Asian *Bradypterus*, all *Locustella*, and *M. pryeri*, is moderately or strongly supported by all analyses except BEST (only complete dataset analysed), and is further supported by one indel, albeit only by two SLAs. Clade B, which includes the African *Bradypterus*, three *Megalurus*, *Dromaeocercus*, *Cincloramphus* and *Eremiornis*, and according to the tree in Fig. 2 also *Schoenicola*, is well supported in all analyses, and receives additional support from one indel, although it is only recovered in two SLAs. These two clades make sense from a biogeographical perspective, as all of the species in clade A breed in the Palearctic or Oriental regions, whereas the species in clade B are Afrotropical/Malagasy (clade E) or Oriental to Australasian (clade F and *M. palustris*) (Bairlein et al., 2006).

Clades C, D, E, F1 and F2 are unanimously well corroborated by the data. The support for the sister relationship between F1 and F2 rests mainly or exclusively on LDH and is lacking in one of the BEST analyses. From a biogeographical and morphological perspective (Bairlein et al., 2006), this is a sensible group (but see comment on *M. palustris*, below). Clades C1–C5 are robust, except for the inclusion of *L. lanceolata* in C5 (see below). However, the relationships among these are uncertain. The relative positions of C1, C2 and C3 vary among analyses. Even the inclusion of these in the same clade is not unanimously well supported, and relies exclusively or mainly on cyt*b*. Also the position of clade C4 rests only or mainly on cyt*b*, and disagrees among different analyses.

B. major is placed in clade C4, as sister to *Locustella naevia*, with good support, in the analyses of the concatenated sequences. However, this should be considered provisional, as it is based on cyt*b* only. Similarly, the inclusion of *L. lanceolata* in clade C5 is tentative. This clade receives high BIC and MLB support, but weak or no statistical support in the species tree or MPB analyses, and is not recovered in any SLA. Also for *B. sylvaticus* and *B. barratti* only cyt*b* is available, and the precise positions within clade E are indeterminate. Regarding the latter, Bairlein et al. (2006) point out that based on similarities in morphology and habitat choice it forms a group with *Bradypterus cinnamomeus*, *Bradypterus lopezi* and *Bradypterus bangwaensis*, in agreement with our results.

The sequences for *S. brevirostris* were acquired from a museum specimen, and no cyt*b* or LDH data were obtained. Due to the missing data, this species was excluded from the species tree analyses. However, BIC places this species as sister to clade E, although with insufficient statistical support, and this position seems reasonable from a biogeographical and morphological perspective (Bairlein et al., 2006). This is further supported by analyses of the nuclear RAG-1 and RAG-2, which place *S. brevirostris* with strong support as sister to a clade with *B. barratti* and *Bradypterus baboecala* (Beresford et al., 2005).

The position of *M. palustris* as sister to *S. brevirostris* plus clade E receives mostly strong support in the different analyses of all loci combined, although this is only inferred by one SLA (cyt*b*). However, this is contradicted by analyses of RAG-1 and RAG-2, according to which *M. palustris* and *C. mathewsi* form a strongly supported clade, which is sister to a clade containing *Schoenicola* and two African *Bradypterus* (Beresford et al., 2005). The tree inferred by the present study is surprising from a morphological and vocal point of view. *M. palustris* resembles the other species of *Megalurus* (clade F) morphologically, whereas it differs in multiple aspects from *S. brevirostris* and the species in clade E (Bairlein et al., 2006). Moreover, the song of *M. palustris* is said to be similar to at least the Philippine populations of *M. timoriensis*, whereas it differs more from *S. brevirostris* and the African *Bradypterus* (Bairlein et al., 2006). In addition, *M. palustris* and the species in clade F are collectively distributed from the Indian Subcontinent via the Philippines and Indonesia to Australia and New Zealand, whereas *S. brevirostris* and the species in clade E occur in the Afrotropics (though the second species of *Schoenicola*, *S. platyurus*, is found in south India). If the position of *M. palustris* inferred here is indeed correct, this implies that the morphological evolution set off in a new direction in the lineage leading to *S. brevirostris*/clade E after these split from a most recent common ancestor with *M. palustris*, whereas the morphological divergence was much more conservative in the lineages leading to *M. palustris* and clade F.

The indels in the nuclear alignments lend further support to the inferred tree. All except one of the nodes with corroborating indel data is unanimously well supported by the different analyses. The exception concerns the most basal node in clade A, which has conflicting inferences.

Drovetski et al. (2004) used mitochondrial ND2 to study the relationships of all *Locustella*, two Asian and three African *Bradypterus* (*B. castaneus*, *B. tacsanowskius*, *B. baboecala*, *B. cinnamomeus*, *B. mariae* [=*B. lopezi mariae*]), and two *Megalurus* (*M. gramineus*, *M. pryeri*). In agreement with our results, they found that Asian *Bradypterus* and *M. pryeri* nested within *Locustella*, and African *Bradypterus* formed a separate clade, as did *M. gramineus*. The relationships within these clades conform with our cyt*b* tree.

4.4. Taxonomic implications

According to our data, the phylogeny strongly disagrees with the current taxonomy at the generic level. We propose a number of taxonomic changes (Fig. 2): (1) that the Asian species of *Bradypterus* and *M. pryeri* be placed in *Locustella*; (2) that *Bradypterus* is restricted to the species in clade E (which includes the type species of this genus, *B. baboecala*), which means that the monotypic genus *Dromaeocercus* is synonymized with *Bradypterus*; (3) that *Schoenicola* is provisionally retained, pending further studies of its affinities based on additional loci and inclusion of the other species of *Schoenicola* (*S. platyurus*, south India) and the two missing African *Bradypterus*; and (4) that *Cincloramphus* and *Eremiornis* are synonymized with *Megalurus*. The last point renders *Megalurus* non-monophyletic, since *M. palustris* is retained in *Megalurus*, and accordingly runs counter to modern taxonomic practice. However, this is a preliminary standpoint, which takes into account the phylogenetic uncertainty with respect to *M. palustris* (conflict between our data, which rest mainly or entirely on cyt*b*, on the one hand, and RAG sequence data, morphology and vocalizations, on the other hand; see above). The alternatives, to treat all of clade B as *Megalurus* (by priority) or to recognize a monotypic *Megalurus* for *palustris* (which is the type species for this genus) and referring to clade F as *Cincloramphus* (by priority), are less appealing at this stage. More data are needed to determine the position of *M. palustris* before this issue can be satisfactorily resolved.

We lack samples of the genera *Amphilais* (monotypic, Madagascar), *Megalurulus* (four species, Melanesia), *Buettikoferella* (monotypic, Timor), and *Chaetornis* (monotypic, Indian Subcontinent), which have been suggested to be closely related to *Megalurus*, and *Elaphrornis* (monotypic, Sri Lanka), which is usually placed in *Bradypterus* (e.g. Bairlein et al., 2006). Future studies will show whether these taxa are part of Locustellidae or not, though in any event they are unlikely to affect the taxonomic changes proposed here.

The taxon *pryeri* has already been suggested to belong in *Locustella* based on morphology (Morioka and Shigeta, 1993), and, as pointed out above, this has been confirmed by mitochondrial ND2 by Drovetski et al. (2004). However, the suggestion by Bairlein et al. (2006) that its two subspecies *pryeri* and *sinensis* might deserve species rank is not corroborated by the present study. The use of the generic name *Bowdleria* for *M. punctatus*, which has been advocated based on osteological characters (Olson, 1990), is not supported by our data.

Some taxa that are currently treated as conspecific appear to be sufficiently divergent (cf. e.g. Hebert et al., 2004; Lovette and Bermingham, 1999; Olsson et al., 2005) in cyt*b* to warrant species status, although this needs to be confirmed by independent data: *L. fasciolata fasciolata* vs. *L. f. amnicola* (also remarked by Drovetski et al., 2004); *B. baboecala tongensis*/*B. b. transvaalensis* vs. *B. b. centralis*/*B. b. elgonensis*; *B. lopezi mariae*/*B. l. usambarae* vs. *B. l. ufipae*; and *M. palustris toklao* vs. *M. p. forbesi*. Also the monotypic *B. luteoventris* needs further study in the light of our results. The two samples of *E. carteri* are surprisingly divergent considering that they are from the same locality, and this needs to be investigated. Conversely, the divergences between the two species pairs *L. pleskei*–*L. ochotensis* and *B. mandelli*–*B. montis* are so slight that their status as separate species need to be studied further. Slight differences between the former pair have previously been found in ND2 (Drovetski et al., 2004), and they have been treated as conspecific (e.g. Williamson, 1968). The two latter have been treated as conspecific (e.g. Watson et al., 1986), but were split based on minor differences in morphology and song (Dickinson et al., 2000).

4.5. Dangers of morphology-based classifications

The present study underscores the well known but still often neglected problem of basing classifications on non-cladistic analyses of morphological characters. The traditional classification of these birds (e.g. Watson et al., 1896; Sibley and Monroe, 1990; Dickinson, 2003; Bairlein et al., 2006) is at variance with the phylogeny presented here regarding every single genus except *Schoenicola* (which comprises only two species, of which only one is included here). These discrepancies result from multiple cases of morphological convergence, e.g. African and Asian "*Bradypterus*", as well as several instances of strongly divergent morphological evolution, e.g. "*Dromaeocercus*", "*Eremiornis*" and "*Cincloramphus*". With regard to the latter genus, Bairlein et al. (2006) state that "striking morphological differences...suggest that the two species may not be closely related and should perhaps be placed in separate genera". The divergence between these is confirmed here, although the phylogeny implies an even more complex pattern of morphological differentiation. Molecular data have previously shown that *Graminicola bengalensis* is not closely related to Locustellidae (Alström et al., 2006; Gelang et al., 2009).

5. Conclusions

For our data, the traditional gene tree methods (Bayesian inference, maximum likelihood, parsimony) and a species tree approach (BEST) yield basically the same topology. In spite of this, we

Taxon	Locality	Sample No./Ref.	GenBank No.				
			Cyt*b*	ODC	Myo	LDH	GAPDH
Abroscopus albogularis fulvifacies	Sichuan, China	DZUG U1932	*HQ706175*	***HQ706303***	***HQ706226***	***HQ706186***	***HQ706264***
Acrocephalus arundinaceus arundinaceus	Austria		*FJ883022*	*FJ883128*	*FJ883098*	*FJ883056*	–
Acrocephalus arundinaceus zarudnyi	Xinjiang, China	NRM 20046787	–	–	–	–	***HQ706300***
Aegithalos caudatus caudatus	Sweden	NRM 976089	*AY228044*	*EU680703*	*AY228281*	***HQ706183***	*FJ357912*
Alauda arvensis arvensis	Sweden	NRM 966614	*AY228047*	*EF625336*	*AY228284*	*HQ333047*	*FJ357913*
Alophoixus pallidus annamensis/khmerensis	C Vietnam	NRM 20046822	*DQ008507*	***HQ706304***	*DQ008559*	–	–
Apalis flavida florisuga	KwaZulu-Natal, South Africa	DZUG U2204; VH B0745 (LDH)	*HQ333036*	*HQ333083*	*HQ333069*	*HQ333049*	*HQ333097*
Bernieria madagascariensis inceleber	Toliara, Madagascar	FMNH 431202	*HQ333038*	*HQ333086*	*HQ333071*	*HQ333052*	*HQ333100*
Bradypterus alishanensis	Taiwan	DZUG U1934	***HQ706133***	***HQ706310***	***HQ706232***	***HQ706192***	***HQ706272***
Bradypterus alishanensis	Taiwan	DZUG U1933	**HQ706132**	–	–	–	–
Bradypterus baboecola tongensis/transvaalensis	Natal, South Africa	NRM 20046782	*DQ008473*	*HQ333084*	*DQ008525*	*HQ333050*	*HQ333098*
Bradypterus baboecala transvaalensis	Gauteng, South Africa	Paulette Bloomer in litt.	AY958221	–	–	–	–
Bradypterus baboecala transvaalensis	Zimbabwe	Paulette Bloomer in litt.	AY958222	–	–	–	–
Bradypterus baboecala tongensis	Malawi	Paulette Bloomer in litt.	958223	–	–	–	–
Bradypterus baboecala tongensis	Zambia	Paulette Bloomer in litt.	958224	–	–	–	–
Bradypterus baboecala centralis	Nigeria	DZUG U1935	***HQ706159***	***HQ706338***	***HQ706259***	***HQ706222***	–
Bradypterus baboecala elgonensis	SW Kenya	VH A0769[a]	FJ883053	–	–	–	–
Bradypterus bangwaensis	Nigeria	DZUG U1025	***HQ706163***	***HQ706330***	***HQ706251***	***HQ706214***	***HQ706292***
Bradypterus carpalis	SW Kenya	VH A0768	***HQ706162***	***HQ706329***	***HQ706250***	***HQ706213***	***HQ706291***
Bradypterus carpalis	Kenya	Paulette Bloomer, in litt.	AY958230	–	–	–	–
Bradypterus castaneus castaneus	S Sulawesi, Indonesia	NRM 20066006	*DQ367925*	***HQ706314***	***HQ706236***	***HQ706196***	***HQ706276***
Bradypterus caudatus unicolor	Mindanao, Philippines	FMNH 392283	***HQ706140***	***HQ706315***	***HQ706237***	***HQ706197***	***HQ706277***
Bradypterus cinnamomeus cinnamomeus	Tanga, Tanzania	ZMUC 121180	–	***HQ706331***	***HQ706252***	***HQ706215***	***HQ706293***
Bradypterus cinnamomeus mildbreadi	Uganda	ZMUC 123143	***HQ706166***	–	–	–	–
Bradypterus cinnamomeus mildbreadi	Uganda	FMNH 355750	AY124541	–	–	–	–
Bradypterus davidi davidi	Hebei, China (m)	NRM 20056595	*DQ367931*	***HQ706316***	***HQ706238***	***HQ706198***	***HQ706278***
Bradypterus davidi davidi	Hebei, China (m)	NRM 20056596	DQ367932	–	–	–	–
Bradypterus davidi davidi	Sichuan, China	ZMUC 117767	DQ367933	–	–	–	–
Bradypterus davidi davidi	Sichuan, China	ZMUC 117768	DQ367934	–	–	–	–
Bradypterus davidi davidi/suschkini	Hong Kong (m)	DZUG U398	**HQ706142**	–	–	–	–
Bradypterus davidi davidi/suschkini	Hong Kong (m)	DZUG U399	**HQ706141**	–	–	–	–
Bradypterus davidi davidi/suschkini	C Mongolia (m)	NRM 20056597	DQ367935	–	–	–	–
Bradypterus graueri	Uganda	DZUG U1937	***HQ706161***	***HQ706328***	***HQ706249***	***HQ706212***	***HQ706290***
Bradypterus kashmirensis	Himachal Pradesh, India	NRM 20056593	*DQ367926*	***HQ706317***	***HQ706239***	***HQ706199***	***HQ706279***
Bradypterus kashmirensis	Himachal Pradesh, India	NRM 20056594	DQ367927	–	–	–	–
Bradypterus lopezi ufipae	Zambia	DZUG U1938	***HQ706165***	***HQ706332***	***HQ706253***	***HQ706216***	***HQ706294***
Bradypterus lopezi mariae	Tanga, Tanzania	ZMUC 05391	**HQ706164**	–	–	–	–
Bradypterus lopezi usambarae	Namuli, Mozambique	Paulette Bloomer, in litt.	AY958227	–	–	–	–
Bradypterus luteoventris	NW Thailand (m)	DZUG U1946	***HQ706144***	***HQ706319***	***HQ706241***	***HQ706201***	***HQ706281***
Bradypterus luteoventris	Tonkin, Vietnam	DZUG U1945	**HQ706143**	–	–	–	–
Bradypterus luteoventris	W Myanmar	DZUG U1944	**HQ706145**	–	–	–	–
Bradypterus major [b]	Ladakh, India	FMNH 240009	**HQ706174**	–	–	–	–
Bradypterus mandelli mandelli	West Bengal, India	DZUG U1339	***HQ706135***	***HQ706311***	***HQ706233***	***HQ706193***	***HQ706273***
Bradypterus mandelli mandelli	NW Thailand	DZUG U1941	**HQ706134**	–	–	–	–
Bradypterus mandelli mandelli/melanorhynchus	Hong Kong (m)	DZUG U1942	**HQ706136**	–	–	–	–
Bradypterus montis	Java	DZUG U1940	***HQ706137***	***HQ706312***	***HQ706234***	***HQ706194***	***HQ706274***

Bradypterus montis	Bali	DZUG U1939	**HQ706138**	–	–	–	–
Bradypterus tacsanowskius	Mongolia	UWBM 57938	*HQ333037*	*HQ333085*	–	*HQ333051*	–
Bradypterus tacsanowskius	Irkutskaya Oblast, Russia	UWBM 51699	**HQ706146**	–	*HQ333070*	–	–
Bradypterus tacsanowskius	Hebei, China (m)	NRM 20046783	DQ008474	–	–	–	*HQ333099*
Bradypterus thoracicus	Sichuan, China	ZMUC 117765	*DQ367929*	***HQ706318***	***HQ706240***	***HQ706200***	***HQ706280***
Bradypterus thoracicus	Sichuan, China	NRM 20056582	DQ367930	–	–	–	–
Bradypterus thoracicus	Qinghai, China	NRM 20056583	DQ367928	–	–	–	–
Bradypterus sylvaticus sylvaticus	George, South Africa	Paulette Bloomer, in litt.	AY958228	–	–	–	–
Bradypterus sylvaticus pondoensis	East London, South Africa	Paulette Bloomer, in litt.	AY958229	–	–	–	–
Cettia cetti cetti	France	DZUG U1936	***HQ706176***	*HQ121555*	***HQ706225***	***HQ706185***	***HQ706263***
Cinclorhamphus cruralis	South Australia	MV B.38407	***HQ706167***	***HQ706334***	***HQ706255***	***HQ706217***	***HQ706296***
Cinclorhamphus mathewsi	Victoria, Australia	MV B.24688	**HQ706169**	–	***HQ706256***	–	–
Cinclorhamphus mathewsi	South Australia	MV B.20019	***HQ706168***	***HQ706335***	–	***HQ706218***	***HQ706297***
Crossleyia xanthophrys	Madagascar	FMNH 393280	***HQ706177***	***HQ706309***	***HQ706231***	***HQ706191***	***HQ706269***
Cryptosylvicola randrianasoloi	Madagascar	FMNH 363849	***HQ706178***	***HQ706308***	***HQ706230***	***HQ706190***	***HQ706268***
Delichon urbicum	Spain	NRM 20046816	*DQ008517*	*EU680721*	*DQ008568*	*HQ333055*	*HQ333103*
Donacobius atricapilla	Paraguay	NRM 966966	*DQ008481*	*EU680723*	*DQ008533*	*HQ333054*	*FJ357915*
Dromaeocercus brunneus	Madagascar	FMNH 384749	***HQ706160***	*EU680724*	*EU680593*	***HQ706211***	***HQ706289***
Eremiornis carteri	W Australia	MV B.24551	***HQ706171***	***HQ706337***	***HQ706258***	***HQ706220***	***HQ706299***
Eremiornis carteri	W Australia	MV B.24554	**HQ706172**	–	–	–	–
Hartertula flavoviridis	Madagascar	FMNH 438721	***HQ706131***	***HQ706307***	***HQ706229***	***HQ706189***	***HQ706267***
Hippolais olivetorum	Kenya	Fregin et al. (2008)	*FJ883048*	*FJ883155*	*FJ883121*	*FJ883080*	–
Hippolais olivetorum	Bulgaria	DZUG U1947	–	–	–	–	***HQ706270***
Hirundo rustica rustica	Sweden	NRM 976238	*DQ008516*	*EF441240*	*AY064258*	–	–
Hirundo rustica rustica	Germany		–	–	–	*HQ333056*	*EF441218*
Iduna similis	Kenya	ZMUC 131329	*FJ899738*	*FJ883159*	*FJ883125*	*FJ883083*	***HQ706271***
Leptopoecile sophiae obscura	Qinghai, China	NRM 20046817	*DQ008518*	*EU680738*	*DQ008569*	***HQ706184***	***HQ706262***
Locustella certhiola ssp.	Hebei, China (m)	NRM 20046785	*DQ008476*	–	*DQ008528*	–	***HQ706286***
Locustella certhiola ssp.	Ningxia, China	DZUG U1388	**HQ706154**	***HQ706325***	–	–	–
Locustella certhiola ssp.	Thailand (m)	DZUG U1284	**HQ706155**	–	–	–	–
Locustella certhiola ssp.	Alakol, Kazakhstan	VH B0756	–	–	–	***HQ706208***	–
Locustella fasciolata amnicola	Sakhalin, Russia	UWBM 47557	***HQ706150***	***HQ706322***	***HQ706244***	***HQ706205***	–
Locustella fasciolata amnicola	Japan	Bernd Leisler, in litt.	Y15689	–	–	–	–
Locustella fasciolata fasciolata	Hebei, China (m)	DZUG U1948	***HQ706151***	***HQ706323***	***HQ706245***	***HQ706206***	***HQ706284***
Locustella fluviatilis	Kenya (m)		AJ004764	–	–	–	–
Locustella fluviatilis	Uncertain	NRM 20026297	–	*HQ121556*	–	–	–
Locustella fluviatilis	Kenya (m)	NRM 20046784	*DQ008475*	–	*DQ008527*	***HQ706203***	*HQ121546*
Locustella lanceolata ssp.[c]	Hebei, China (m)	DZUG U1949	***HQ706139***	***HQ706313***	***HQ706235***	***HQ706195***	***HQ706275***
Locustella lanceolata ssp.	China (m?)	–	DQ119524	–	–	–	–
Locustella lanceolata ssp.	China (m?)	–	DQ119525	–	–	–	–
Locustella luscinioides luscinioides	Sweden	NRM 20056589	***HQ706149***	***HQ706321***	***HQ706243***	***HQ706204***	***HQ706283***
Locustella luscinioides luscinioides	Germany	–	AJ004763	–	–	–	–
Locustella luscinioides ssp.	Israel (m)	DZUG U1950	**HQ706148**	–	–	–	–
Locustella naevia naevia	Sweden	DZUG U1951	***HQ706147***	***HQ706320***	***HQ706242***	***HQ706202***	***HQ706282***
Locustella ochotensis	Philippines (m)	DZUG U1619	***HQ706156***	***HQ706326***	–	–	–
Locustella ochotensis	Philippines (m)	DZUG U1621	–	–	***HQ706247***	–	***HQ706287***
Locustella ochotensis	Usuria, Russia	VH A0694	–	–	–	***HQ706209***	–
Locustella ochotensis	Taiwan (m)	DZUG 2101	***HQ706157***	–	–	–	–

(*continued on next page*)

a (*continued*)

Taxon	Locality	Sample No./Ref.	GenBank No.				
			Cyt*b*	ODC	Myo	LDH	GAPDH
Locustella pleskei	Izu isl., Japan	DZUG U1953	–	***HQ706327***	***HQ706248***	***HQ706210***	***HQ706288***
Locustella pleskei	Izu isl., Japan	DZUG U1952	***HQ706158***	–	–	–	–
Locustella pleskei	Izu isl., Japan	Takema Saitoh, in litt.	AB159188	–	–	–	–
Locustella pleskei	Deogu-do isl., South Korea	Takema Saitoh, in litt.	AB159191	–	–	–	–
Megalurus gramineus goulburni	South Australia, Australia	ANWC D224	*HQ333042*	*HQ333091*	*HQ333074*	*HQ333060*	*HQ333108*
Megalurus palustris toklao	Punjab, India	NRM 20046786	*DQ008477*	*EU680741*	*DQ008529*	***HQ706221***	*FJ357917*
Megalurus palustris forbesi	Negros, Philippines	ZMUC 02031	FJ883052	FJ883161	–	FJ883089	–
Megalurus pryeri pryeri	Japan	DZUG U1954	–	***HQ706324***	***HQ706246***	***HQ706207***	***HQ706285***
Megalurus pryeri pryeri	Japan	DZUG U1955	***HQ706152***	–	–	–	–
Megalurus pryeri pryeri	Japan	Bernd Leisler, in litt.	AJ004321	–	–	–	–
Megalurus pryeri pryeri	Japan	Bernd Leisler, in litt.	AJ004322	–	–	–	–
Megalurus pryeri pryeri	Japan	Bernd Leisler, in litt.	AJ004323	–	–	–	–
Megalurus pryeri sinensis	China (status unknown)	–	EU016090	–	–	–	–
Megalurus pryeri sinensis	China (status unknown)	–	EU016091	–	–	–	–
Megalurus pryeri sinensis	Jiangxi, China (m)	DZUG U1956	**HQ706153**	–	–	–	–
Megalurus punctatus vealeae	New Zealand	AWMM B.10962	**HQ706173**	–	–	–	–
Megalurus timoriensis	Luzon, Philippines	ZMUC 119529	***HQ706170***	***HQ706336***	***HQ706257***	***HQ706219***	***HQ706298***
Melocichla mentalis mentalis	Nigeria	NRM 20046804	*DQ008500*	*HQ333090*	*DQ008551*	–	*HQ333107*
Melocichla mentalis mentalis	Ivory Coast	VH A1550	–	–	–	*HQ333059*	–
Mirafra javanica williamsoni	Thailand	NRM 20046819	*DQ008520*	*HQ333089*	*DQ008571*	*HQ333058*	*HQ333106*
Orthotomus sutorius inexpectatus	NW Thailand	NRM 20046795	*DQ008491*	*HQ333092*	*DQ008542*	–	*HQ333109*
Orthotomus sutorius guzuratus	NW India	VH A1581	–	–	–	*HQ333061*	–
Oxylabes madagascariensis	Madagascar	FMNH 438719	***HQ706179***	***HQ706306***	***HQ706228***	***HQ706188***	***HQ706266***
Phylloscopus sindianus lorenzii	NE Turkey	DZUG U1957	***HQ706180***	***HQ706340***	–	–	–
Phylloscopus sindianus lorenzii	NE Turkey	DZUG U1958	–	–	***HQ706261***	–	–
Phylloscopus sindianus lorenzii	Caucasus	VH B0799	–	–	–	***HQ706224***	–
Phylloscopus sindianus sindianus	Pakistan	DZUG U1959	–	–	–	–	***HQ706302***
Pycnonotus barbatus inornatus	Mauretania	DZUG U2047	*HQ333043*	*HQ333093*	*HQ333075*	*HQ333062*	*HQ333110*
Schoenicola brevirostris alexinae	Kenya	NRM 569624	–	***HQ706333***	***HQ706254***	–	***HQ706295***
Seicercus tephrocephalus	W Myanmar	DZUG U1960	***HQ706182***	***HQ706339***	***HQ706260***	–	***HQ706301***
Seicercus tephrocephalus	W Myanmar	DZUG U1961	–	–	–	***HQ706223***	–
Stachyris nigriceps yunnanensis/rileyi	Ha Tinh province, C Vietnam	NRM 20026627	*HQ333045*	*HQ333095*	–	*HQ333065*	*HQ333112*
Stachyris nigriceps yunnanensis	Tonkin, N Vietnam[d]	NRM 947308	–	–	*AY228321*	–	–
Sylvia atricapilla atricapilla	Sweden	NRM 976380	–	*EF441254*	*AY887727*	–	*EF441232*
Sylvia atricapilla atricapilla	Germany	–	*Z73494*	–	–	–	–
Sylvia atricapilla atricapilla	Germany	VH A0364	–	–	–	*HQ333067*	–
Thamnornis chloropetoides	Madagascar	FMNH 436448	*HQ333046*	*HQ333096*	*HQ333077*	*HQ333068*	*FJ357923*
Xanthomixis apperti	Madagascar	FMNH 427370	***HQ706181***	***HQ706305***	***HQ706227***	***HQ706187***	***HQ706265***

[a] Same sample as FJ883162.
[b] Sequence obtained from Trevor Price and Udayan Borthakur.
[c] On geographical grounds, most likely nominate subspecies, but samples collected during migration, so subspecies *hendersonii* cannot be eliminated.
[d] According to Peter Nilsson (in litt.) (not given in published paper).

consider the latter to be a step forward, since it accounts for the ubiquitous heterogeneity in gene trees, thereby providing more realistic support than concatenation for nodes with incongruence among loci. However, BEST is computationally intense, and convergence proved difficult to attain, even in extremely long runs (up to 1×10^9 generations). We suggest that if other analyses indicate the presence of some well corroborated monophyletic subgroups, analysing these separately is likely to help BEST converge more quickly.

The phylogeny strongly disagrees with the current taxonomy at the generic level. We propose a revised classification that recognizes four instead of seven genera. One of these (*Megalurus*) is actually non-monophyletic according to our data, but we stress that this classification is tentative and takes account of the phylogenetic uncertainty (i.e. conflict between our results, which in this case rest on only one locus, and previously published data based on another locus, as well as morphology and vocalizations). Analysis of multiple cyt*b* haplotypes, including several different subspecies of polytypic species, suggests several cases where taxonomic revision is warranted.

Acknowledgements

We are most grateful to the following institutions or persons for samples: Göran Frisk, Ulf Johansson and Peter Nilsson and the Swedish Museum of Natural History, Stockholm, Sweden; Jon Fjeldså and Jan Bolding Kristensen and the Zoological Museum of the University of Copenhagen, Copenhagen, Denmark; David Willard and the Field Museum of Natural History, Chicago, USA; Museum Victoria, Melbourne, Australia; Brian Gill and the Auckland War Memorial Museum, Auckland, New Zealand; Australian National Wildlife Collections (CSIRO); Sharon Birks and the University of Washington Burke Museum, Seattle, USA; Martin Haase and Vogelwarte Hiddensee, Zoological Institute and Museum, Ernst Moritz Arndt University of Greifswald, Greifswald, Germany; Staffan Andersson; Staffan Bensch; Geoff Carey and Paul Leader; Ross McGregor; Magnus Jäderblad; Ulf Ottosson; Bo Petersson; Patrik Rhönnstad; Phil Round; Yoshimitsu Shigeta; Johan Ställberg; Trevor Price; Mats Waern; and Anders Waldenström. P.A. is also most grateful to Cheng-te Yao and the Taiwan Endemic Species Research Institute for invaluable assistance in the field. Dario Zuccon kindly sequenced *S. brevirostris*; Trevor Price and Udayan Borthakur generously provided a cytochrome *b* sequence of *B. major*; and Paulette Bloomer, Bernd Leisler and Takema Saitoh kindly provided information on geographic origin of some cytochrome *b* sequences in GenBank. Scott Edwards is gratefully acknowledged for his most valuable discussions and suggestions on BEST analyses, and for running three of the shorter BEST analyses for us; Normand David for comments on the endings of the scientific names proposed here; Edward Dickinson and Lars Larsson for various information; Peter Kennerley for comments on the Locustellidae introduction; and two anonymous referees for constructive criticism. P.A. also gratefully acknowledges the Riksmusei Vänners Linnaeus award, which has allowed him to devote time to this study. The Swedish Research Council provided financial support (Grant No. 621-2006-3194 to U.O. and 621-2007-5280 to P.E.).

Appendix A

List of samples (in alphabetical order), with GenBank accession numbers. Taxonomy follows Dickinson (2003), except for splitting of *Bradypterus davidi* and *B. kashmirensis* from *B. thoracicus* (Alström et al., 2008), inclusion of the recently described *Bradypterus alishanensis* (Rasmussen et al., 2000), and *Iduna similis* being moved from genus *Chloropeta* (Fregin et al., 2009). ANWC = Australian National Wildlife Collection (CSIRO), Canberra, Australia; AWMM = Auckland War Memorial Museum, Aukland, New Zealand; DZUG = Department of Zoology, University of Gothenburg, Göteborg, Sweden; FMNH = Field Museum of Natural History, Chicago, USA; MNHN = Muséum National d'Histoire Naturelle, Paris, France; MV = Museum Victoria, Melbourne, Australia; NRM = Swedish Museum of Natural History, Stockholm, Sweden; UWBM = University of Washington Burke Museum, Seattle, USA; VH = Vogelwarte Hiddensee, Zoological Institute and Museum, Ernst Moritz Arndt University of Greifswald, Greifswald, Germany; ZMUC = Zoological Museum of the University of Copenhagen, Copenhagen, Denmark. m = Sample collected on migration or in winter quarters. Sequences that are new to this study are in bold, and sequences included in multilocus analyses are in italics.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.12.012.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans Auto Control. 19, 716–723.
Allen, E.S., Omland, K.E., 2003. Novel intron phylogeny supports plumage convergence in orioles (*Icterus*). Auk 120, 961–969.
Alström, P., Ericson, P.G.P., Olsson, U., Sundberg, P., 2006. Phylogeny and classification of the avian superfamily Sylvioidea. Mol. Phylogenet. Evol. 38, 381–397.
Alström, P., Rasmussen, P.C., Olsson, U., Sundberg, P., 2008. Species delimitation based on multiple criteria: the Spotted Bush Warbler *Bradypterus thoracicus* complex (Aves, Megaluridae). Zool. J. Linn. Soc. 154, 291–307.
Bairlein, F., Alström, P., Aymí, R., Clement, P., Dyrcz, A., Gargallo, G., Hawkins, F., Madge, S., Pearson, D., Svensson, L., 2006. Family Sylviidae (warblers). In: del Hoyo, J., Elliott, A., Christie, D.A. (Eds.), Handbook of the Birds of the World, vol. 12. Lynx Edicions, Barcelona., pp. 492–709.
Belfiore, N.M., Liu, L., Moritz, C., 2008. Multilocus phylogenetics of a rapid radiation in the genus Thomomys (Rodentia: Geomyidae). Syst. Biol. 57, 294–310.
Beresford, P., Barker, F.K., Ryan, P.G., Crowe, T.M., 2005. African endemics span the tree of songbirds (Passeri): molecular systematics of several evolutionary 'enigmas'. Proc. Roy. Soc. Lond., B. 272, 849–858.
Bock, W.J., 1994. History and nomenclature of avian family-group names. Bull. Am. Mus. Nat. Hist. 222, 1–283.
Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56, 643–655.
Brumfield, R.T., Liu, L., Lum, D.E., Edwards, S.V., 2008. Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae, *Manacus*) from multilocus sequence data. Syst. Biol. 57, 719–731.
Cranston, C., Hurwitz, B., Ware, D., Stein, L., Wing, R.A., 2009. Species trees from highly incongruent gene trees in rice. Syst. Biol. 58, 489–500.
Dickinson, E.C., 2003. The Howard and Moore Complete Checklist of the Birds of the World. Christopher Helm, London.
Dickinson, E.C., Rasmussen, P.C., Round, P.D., Rozendaal, F.G., 2000. Systematic notes on Asian birds. 1. A review of the russet bush-warbler *Bradypterus seebohmi* (Ogilvie-Grant, 1895). Zool. Verh. Leiden 331, 12–64.
Drovetski, S.V., Zink, R.M., Fadeev, I.V., Nesterov, E.V., Koblik, E.A., Red'kin, Y.A., Rohwer, S., 2004. Mitochondrial phylogeny of *Locustella* and related genera. J. Avian Biol. 35, 105–110.
Eckert, A.J., Carstens, B.C., 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. Mol. Phylogenet. Evol. 49, 832–842.
Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.
Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA 104, 5936–5941.
Fjeldså, J., Zuccon, D., Irestedt, M., Johansson, U.S., Ericson, P.G.P., 2003. *Sapayoa aenigma*: a new world representative of 'Old World suboscines'. Proc. Roy. Soc. Lond. B 270 (Suppl.), 238–241.
Fregin, S., Haase, M., Olsson, U., Alström, P., 2009. Multi-locus phylogeny of the family *Acrocephalidae* (Aves: Passeriformes) – the traditional taxonomy overthrown. Mol. Phylogenet. Evol. 52, 866–878.
Gelang, M., Cibois, A., Pasquet, E., Olsson, U., Alström, P., Ericson, P.G.P., 2009. Phylogeny of babblers (Aves, Passeriformes): major lineages, family limits and classification. Zool. Scr. 35, 225–236.
Gu, X., Fu, Y.X., Li, W.H., 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol. Biol. Evol. 12, 546–557.
Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of birds through DNA barcodes. PLoS Biol. 2, e312. doi:10.1371/journal.pbio.0020312.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
Huelsenbeck, J.P., Ronquist, F., 2005. MRBAYES version 3.1.2. <http://mrbayes.scs.fsu.edu/download.php>.
Irestedt, M., Ohlson, J.I., Zuccon, D., Källersjö, M., Ericson, P.G.P., 2006. Nuclear DNA from old collections of avian study skins reveals the evolutionary history of the Old World suboscines (Aves, Passeriformes). Zool. Scr. 35, 567–580.
Johansson, U., Fjeldså, J., Bowie, R.C.K., 2008. Phylogenetic relationships within Passerida (Aves: Passeriformes): a review and a new molecular phylogeny based on three nuclear intron markers. Mol. Phylogenet. Evol. 48, 858–876.
Jobb, G., 2008. TREEFINDER (version of October 2008). <http://www.treefinder.de>.
Jobb, G., von Haeseler, A., Strimmer, K., 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4, 18. doi:10.1186/1471-2148-4-18.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
Lanave, C., Preparata, C., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86–93.
Leaché, A.D., 2009. Species tree discordance traces to phylogeographic clade boundaries in North American fence lizards (*Sceloporus*). Syst. Biol. 58, 547–559.
Linnen, C.R., Farrell, B.D., 2008. Comparison of methods for species-tree inference in the sawfly genus *Neodiprion* (Hymenoptera: Diprionidae). Syst. Biol. 57, 876–890.
Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542–2543.
Liu, L., Edwards, S.V., 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58, 452–460.
Liu, L., Pearl, D.K., 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56, 504–514.
Liu, L., Pearl, D.K., Brumfield, R., Edwards, S.V., 2008. Estimating species trees using multiple-allele DNA sequence data. Evolution 62, 2080–2091.
Lovette, I.J., Bermingham, E., 1999. Explosive speciation in the New World *Dendroica* warblers. Proc. Roy. Soc. Lond. B 266, 1629–1636.
McGuire, J.A., Witt, C.C., Altshuler, D.L., Remsen Jr., J.V., 2007. Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. Syst. Biol. 56, 837–856.
Morioka, H., Shigeta, Y., 1993. Generic allocation of the japanese Marsh Warbler *Megalurus pryeri* (Aves: Sylviidae). Bull. Nat. Sci. Mus. Tokyo A 19, 37–43 (in Japanese).
Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. J. Roy. Stat. Soc., B 56, 3–48.
Nylander, J.A.A., 2004. MRMODELTEST2 version 2.2. <http://www.abc.se/~nylander>.
Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.
Olson, S.L., 1990. Osteology and systematics of the fernbirds (*Bowdleria*; Sylviidae). Notornis 37, 161–171.
Olsson, U., Alström, P., Ericson, P.G.P., Sundberg, P., 2005. Non-monophyletic taxa and cryptic species – evidence from a molecular phylogeny of leaf-warblers (Phylloscopus, Aves). Mol. Phylogenet. Evol. 36, 261–276.
Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Rambaut, A., Drummond, A.J., 2009. Tracer version 1.5.0. <http://beast.bio.ed.ac.uk>.
Rasmussen, P.C., Round, P.D., Dickinson, E.C., Lambert, F.G., 2000. A new bush-warbler (Sylviidae, Bradypterus) from Taiwan. Auk 117, 279–289.
Rodríguez, J., Oliver, L., Marín, A., Medina, R., 1990. The general stochastic model of nucleotide substitution. J. Theo. Biol. 142, 485–501.
Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Sibley, C.G., Ahlquist, J.E., 1990. Phylogeny and Classification of Birds. A Study in Molecular Evolution. Yale University Press, New Haven and London.
Sibley, C.G., Monroe Jr., B.L., 1990. Distribution and Taxonomy of Birds of the World. Yale University Press, New Haven.
Sorensen, M.D., Quinn, T.W., 1998. Numts: a challenge for avian systematics and population biology. Auk 115, 214–221.
Swofford, D.L., 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods) version 4.0b10. Sinauer Associates, Sunderland, MA.
Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lect. Mater. Life Sci. 17, 57–86.
Thomson, R.C., Shedlock, A.M., Edwards, S.V., Shaffer, H.B., 2008. Developing markers for multilocus phylogenetics in non-model organisms: a test case with turtles. Mol. Phylogenet. Evol. 49, 514–525.
Watson, G.E., Traylor Jr., M.A., Mayr, E., 1986. In: Mayr, E., Cottrell, G.W. (Eds.), Checklist of Birds of the World, vol. 11. Museum of Comparative Zoology, Cambridge, Massachusetts.
Williamson, K., 1968. Identification for Ringers. The Genera *Cettia*, *Locustella*, *Acrocephalus* and *Hippolais*, third ed. British Trust for Ornithology, Tring.
Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39, 306–314.

Molecular Phylogenetics and Evolution 62 (2012) 319–328

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Pitfalls in comparisons of genetic distances: A case study of the avian family Acrocephalidae

Silke Fregin [a,*], Martin Haase [a], Urban Olsson [b], Per Alström [c]

[a] Vogelwarte Hiddensee, Zoological Institute and Museum, Ernst Moritz Arndt University of Greifswald, 17489 Greifswald, Germany
[b] Department of Zoology, University of Göteborg, Box 463, SE-405 30 Göteborg, Sweden
[c] Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, SE-750 07 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 21 January 2011
Revised 9 August 2011
Accepted 3 October 2011
Available online 17 October 2011

Keywords:
Acrocephalidae
Barcoding
Cytochrome *b*
DNA taxonomy
Genetic distance
Substitution model

ABSTRACT

Genetic distances are increasingly being used for identification and species delimitation, especially since the introduction of "barcoding". While for phylogenetic inferences great care is generally taken to choose the best-fit evolutionary model, this is usually neglected in calculating genetic distances. Moreover, distances obtained from others than best-fit models, different lengths of sequences, and even different loci are often freely compared. We examined the influence of different methods on calculating genetic distances using mitochondrial cytochrome *b* sequences for the passerine family Acrocephalidae.

We found substantial differences between: (1) corrected distances based on the best-fit model (TrN + Γ) vs. uncorrected *p*-distances; (2) distances calculated based on different parts of the same gene; and (3) distances calculated using the methods of "complete deletion" vs. "pairwise deletion" for sequences that included uncertain nucleotides. All these methodological differences affected comparisons between species and potential taxonomical conclusions.

We suggest that (1) different loci are incomparable. (2) Only perfectly homologous regions (same length, same part of locus) should be compared. (3) In the case of sequences with some uncertain nucleotides, only distances calculated by the method of "complete deletion" are fully comparable. (4) Only distances based on the optimal substitution model should be used. (5) Even within the same locus, corrected genetic distances are unique to the study in which they are calculated, as they are conditional on the particular dataset and model selected for that dataset.

1. Introduction

Genetic distances have frequently been used in assessments of species status of closely related taxa and in a diversity of evolutionary studies (e.g. Bradley and Baker, 2001; Burbrink et al., 2000; Cagnon et al., 2004; Hickerson et al., 2003; Hung et al., 1999; Kergoat et al., 2011; Klicka et al., 1999; Macey et al., 2001; Olsson et al., 2005; Parkin et al., 2004; Petren et al., 2005; Price, 2010; Wesson et al., 1993). Evaluation of species status has become even more widespread after the introduction of "barcoding" (Hajibabaei et al., 2006; Hebert et al., 2003a,b, 2004; Hunt et al., 2010; Johnsen et al., 2010; Pons et al., 2006; Semina et al., 2007; Vences et al., 2005). Based on a broad study of partial cytochrome oxidase subunit I (COX1 or CO1) "barcodes" of North American birds, Hebert et al. (2004) concluded that divergences between species (Kimura-two-parameter corrected; Kimura, 1980) are usually ≥10 times higher than mean intraspecific differences. They suggested that this "barcoding gap" could be used as a screening tool for the identification of species. Another study (Hebert et al., 2003b) found that mean intraspecific CO1 distances were generally <2%, and in the Barcoding of Life Data System (BOLD) sequences up to a maximum divergence of 2% are considered to be conspecific, whereas their identification system for unknown sequences works with sequence divergences <1% (Ratnasingham and Hebert, 2007). Many studies have confirmed the "barcoding gap", also for other genetic markers than CO1 (e.g. Aliabadian et al., 2009; Barrett and Hebert, 2005; Dalebout et al., 2007; Derycke et al., 2010; Hubert et al., 2008; Kerr et al., 2007; Nijman and Aliabadian, 2010), while others have criticized this concept for various reasons. Mostly, no barcoding gap was found (e.g. Davison et al., 2009; Meier et al., 2006; Langhoff et al., 2009). Wiemers and Fiedler (2007) and Meier et al. (2008) found barcoding gaps to be artifacts of insufficient sampling and of the incorrect use of mean distances instead of smallest distances, respectively. Moreover, Meyer and Paulay (2005) concluded based on densely sampled species that approximately 20% of the species will be overlooked using barcodes. Numerous

* Corresponding author. Fax: +49 3834 864332.
E-mail addresses: silke.fregin@uni-greifswald.de (S. Fregin), martin.haase@uni-greifswald.de (M. Haase), urban.olsson@zool.gu.se (U. Olsson), per.alstrom@artdata.slu.se (P. Alström).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.10.003

studies have reported notable non-correspondence between taxonomic treatment and genetic divergence. These discrepancies have been suggested to be possibly the result of introgression (e.g. Alström et al., 2008a; Irwin et al., 2009; Olsson et al., 2010; Weckstein et al., 2001), incomplete lineage sorting (e.g. Alexander et al., 2009; Langhoff et al., 2009), recent speciation (e.g. Besansky et al., 2003; Haase, 2005; Haase et al., 2007; Hebert et al., 2010) or improper taxonomic treatment (e.g. Alström et al., 2008b; Gomez et al., 2002; Helbig et al., 1995, 1996; Martens et al., 2004; Olsson et al., 2005; Wilcox et al., 1997). In summary, there is no universal cut-off value for the delimitation of species, and many authors have repeatedly argued that in the absence of other data, levels of genetic divergence are inadequate for the definition of species across a variety of taxa (e.g. Burns and Naoki, 2004; Cognato, 2006; Dalebout et al., 2007; Davis and Nixon, 1992; Hajibabaei et al., 2006; Kerr et al., 2009a; Meyer and Paulay, 2005; Sangster, 2000; Roe and Sperling, 2007; Rubinoff et al., 2006; Whitworth et al., 2007; Wiemers and Fiedler, 2007).

Using genetic distances as yardsticks for species limits is not only a problematic issue regarding cut-off values. A more basic question concerns the correction of distances to account for multiple substitutions at certain sites (Swofford et al., 1996). In phylogenetic analyses, model selection is considered important as the substitution model always influences branch lengths and may consequently affect the tree topology (Johnson and Omland, 2003; Lemmon and Moriarty, 2004; Posada and Crandall, 2001; Posada and Buckley, 2004; Posada, 2008; Yang et al., 1994). The improper use of uncorrected as well as under-corrected distances (e.g. Benzoni et al., 2010; Loader et al., 2010; Palma et al., 2010) will lead to underestimation of the actual differences between long separated taxa (Arbogast et al., 2002).

Another basic issue concerns the comparability of genetic distances. It is well known that different loci have different mean rates of evolution, and accordingly levels of divergence between taxa depend on the loci being compared (Aliabadian et al., 2009; Broughton and Reneau, 2006; Lin and Danforth, 2004; Nijman and Aliabadian, 2010; Palma et al., 2010). However, this is often neglected, as studies based on different loci are frequently indiscriminately compared (e.g. Baker et al., 2003; Helbig et al., 1995; Johnson and Cicero, 2004; Zhang et al., 2007). Also different parts of the same gene are known to have different evolutionary rates (Cicero and Johnson, 1995; Galtier et al., 2006; Griffiths, 1997; Klicka et al., 2001), and this may confound comparisons based on different gene fragments, or overlapping fragments of unequal lengths even when they are from the same locus.

The aims of the present study are to examine how genetic distances are affected by (1) different correction methods; (2) the use of different parts of the same locus under the same correction model; and (3) for sequences including uncertain nucleotides, the choice of "complete deletion" (an unknown or uncertain nucleotide N at any position in the alignment results in deletion of this position over the whole alignment) or "pairwise deletion" (an unknown or uncertain nucleotide N is only deleted for the pair of species under comparison). For this study we used a dataset comprising cytochrome *b* (cyt *b*) sequences for most species and a large number of subspecies from the avian family Acrocephalidae. This is a suitable group for a study of genetic distances, as it is densely sampled and contains taxa at different degrees of divergence, from old species to recently separated subspecies.

The avian warbler family Acrocephalidae (sensu Alström et al., 2006; Johansson et al., 2008) comprises the genera *Acrocephalus*, *Hippolais*, *Iduna*, *Calamonastides* and *Nesillas* (sensu Fregin et al., 2009). They breed widely across the Old World and Australasia. Most members of this family occur in wet habitats, such as reedbeds and other marshland vegetation, but some, e.g. all *Hippolais*, *Iduna* and *Nesillas* species, live in drier habitats (Bairlein et al., 2006). Many species are inconspicuous, and their songs are generally the best identification feature for closely related species (Cramp et al., 1992; Bairlein et al., 2006).

2. Methods

2.1. Material and laboratory procedures

We obtained blood, feathers or muscle tissue from 33 individuals of 15 species of *Acrocephalus*, *Iduna* and *Hippolais*. In addition, we downloaded 87 cyt *b* sequences from 41 species from GenBank. All of the new sequences used in the present study come from species that have been studied previously (e.g. Cibois et al., 2007; Fregin et al., 2009; Helbig and Seibold, 1999; Leisler et al., 1997) and all new haplotypes closely agree with sequences available on GenBank. In total, we obtained sequences from 120 individuals belonging to 61 taxa, including all species of *Acrocephalus* except one Eurasian (*Acrocephalus sorghophilus*), one Indian Ocean (*Acrocephalus rodericanus*), and four Pacific island ones (*Acrocephalus luscinius*, *Acrocephalus rehsei*, *Acrocephalus syrinx*, *Acrocephalus vaughani*), and all species of *Calamonastides*, *Hippolais* and *Iduna* (sensu Fregin et al., 2009). For laboratory procedures see Olsson et al. (2005) and Fregin et al. (2009).

The more distantly related *Nesillas* was not included due to a shortage of sequence data. The sequences used are given in Appendix.

Species names follow Bairlein et al. (2006), except for the changes of some generic names proposed in Fregin et al. (2009) and the species names of some Polynesian and Australian species proposed by Cibois et al. (2007). Subgeneric names used in Helbig and Seibold (1999) and Leisler et al. (1997) were adopted here.

2.2. Phylogenetic and distance analysis

Sequences were aligned by eye in BioEdit (Hall, 1999). Identical sequences were only included if they were from different subspecies. According to AIC (Akaike, 1974), jModelTest (Posada, 2008) determined the Tamura–Nei model (TrN, Tamura and Nei, 1993), assuming rate variation across sites according to a discrete gamma distribution with four rate categories (Γ; Yang et al., 1994) as the best fitting substitution model.

Genetic distances, uncorrected *p*-distances and TrN + Γ corrected distances, were calculated in PAUP* (Swofford, 2003). If not otherwise stated, genetic distances are given in the following order: uncorrected/corrected. To compare the effects of the two differentways of handling missing data, i.e. "pairwise deletion" and "complete deletion", respectively (see Section 1 for explanation), an alignment with all positions containing at least one unknown/ambiguous nucleotide removed which corresponds to "complete deletion", was produced manually. This was done, as PAUP* uses pairwise deletion, and offers no choice between the two methods. "Complete deletion" was used for the comparison of genetic distances based on the first and second halves, respectively, of our alignment. Sequence statistics were calculated in DNAsp 5.10 (Librado and Rozas, 2009). For better clarity we restrict detailed comparisons to the subgenus *Calamodus* (sensu Helbig and Seibold, 1999), the small striped warblers. Complete distance matrices are available in the electronic Supplement.

3. Results

3.1. Sequence characteristics

The entire cyt *b* alignment comprised 879 nucleotides, including some undetermined nucleotides at certain positions for some

sequences, which were the basis for comparison of the two different deletion methods. It should be noted that due to these uncertainties, the number of nucleotides compared varied between different pairwise comparisons. After deletion of all uncertain positions (''complete deletion''), only 718 nucleotides were effectively usable.

3.2. Comparison of uncorrected vs. corrected genetic distances

For comparison of uncorrected and corrected (TrN + Γ) genetic distances the method of ''complete deletion'' was used. In general, corrected distances were markedly higher than uncorrected distances, and the difference between these methods was considerably greater in comparisons of distantly related taxa than in comparisons of closely related taxa (Supplement 1). Between genera, uncorrected distances ranged from 0.060 to 0.118, while corrected distances varied from 0.085 to 0.230 (Fig. 1). In only one case (*Acrocephalus* vs. *Iduna*) did ranges of genetic distances of both methods overlap to a greater extent.

Within genera, corrected distances (0.003–0.222) were generally higher than uncorrected ones (0.003–0.111), but the differences were less pronounced than between genera (Fig. 2). The lowest genetic distances in *Hippolais*, *Iduna* and *Calamodus* were much higher than in the rest of the family.

Applying the 2% threshold for conspecific sequences proposed by BOLD (Barcoding of Life Data System, Ratnasingham and Hebert, 2007), 76/65 pairwise sequence comparisons involving 14/14 species currently treated as separate species were lower than 2% when uncorrected/corrected distances were used. This concerned mainly species from Pacific islands and the sister pairs *Acrocephalus brevipennis*/*Acrocephalus rufescens* and *Acrocephalus baeticatus*/*Acrocephalus scirpaceus*.

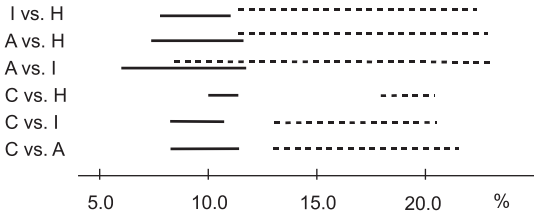

Fig. 1. Comparison of genetic distances between genera. Unbroken line: uncorrected *p*-distances: dotted line: TrN + Γ corrected distances. Abbreviations: A: *Acrocephalus*, C: *Calamonastides*, H: *Hippolais*, I: *Iduna*.

3.3. Comparison of ''pairwise deletion'' vs. ''complete deletion''

For *Calamodus*, uncorrected and TrN + Γ corrected distances, both calculated under ''complete deletion'' and ''pairwise deletion'', respectively, are shown in Table 1. As is evident, the choice of deletion method strongly affects the genetic distances. For example, the two haplotypes of *Acrocephalus bistrigiceps* (1 and 2) had no missing nucleotides in our alignment of 879 nucleotides, thus ''pairwise deletion'' used the full length and calculated the ''true'' genetic distance between these sequences 0.031/0.034 (uncorrected/corrected). In contrast, the use of ''complete deletion'', whereby the sequences were trimmed to 719 nucleotides, reduced the genetic distance to 0.019/0.020. In the entire data set, 65 pairwise comparisons involving 14 species were lower than 2% under the ''complete deletion'' model, but only 37 pairwise comparisons involving 10 species using ''pairwise deletion''. However, several cases, e.g. *Acrocephalus paludicola* (4) vs. *Acrocephalus schoenobaenus* (12), were hardly affected by the choice of method (see also Appendix and electronic supplement).

3.4. Comparison of different parts of cyt b

The first half (359 nucleotides) and second half (359 nucleotides) of the ''complete deletion'' alignment of all taxa did not differ much regarding their sequence statistics: the first half had 134 parsimony informative of 151 variable sites and the second half had 131 parsimony informative of 152 variable sites. Both halves contained nine fourfold degenerated sites, where all four nucleotides occur at the same position in the alignment. Based on the first half, 81 pairwise sequence comparisons concerning 14 species were below 2%. In contrast, based on the second half 177 comparisons involving 18 species failed to reach the 2% threshold. This reflects the general pattern of higher genetic distances calculated from the first half of cyt *b* (U-test: $p < 0.001$). However, this picture changed when comparisons were restricted to subgroups, e.g., in *Calamodus* genetic distances were higher in the second half of cyt *b* (Table 2), although these differences were statistically not significant (U-tests, $p > 0.19$).

One of the largest changes occurred between *A. paludicola* (4) and *A. schoenobaenus* (8): 0.081/0.056 (first half, uncorrected/corrected) vs. 0.111/0.071 (second half). Large differences were also noted within species, e.g. in *Acrocephalus melanopogon mimicus*, where genetic distances increased from 0.011/0.010 (first half) to 0.025/0.027 (second half). In a few cases there was no change at all [e.g. *A. schoenobaenus* (8) vs. *A. schoenobaenus* (9)].

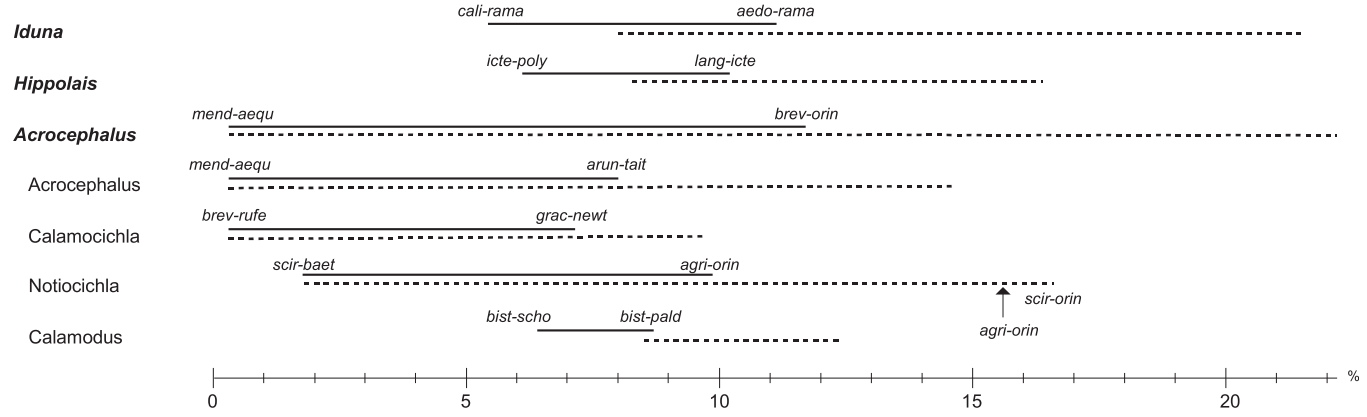

Fig. 2. Comparison of genetic distances within Acrocephalidae sub-/genera. Unbroken line: uncorrected, dotted line: TrN + Γ corrected distances. Genera are in bold, while subgenera are in regular font. Names at the beginning and end of the unbroken lines indicate species pairs with lowest and highest values, respectively. These are identical for dotted lines if not otherwise stated. Abbreviations in alphabetical order: aedo: *aedon*, aequ: *aequinoctialis*, agri: *agricola*, arun: *arundinaceus*, avic: *avicenniae*, baet: *baeticatus*, bist: *bistrigiceps*, brev: *brevipennis*, cali: *caligata*, grac: *gracilirostris*, gris: *griseldis*, icte: *icterina*, mend: *mendanae*, newt: *newtoni*, oliv: *olivetorum*, orin: *orinus*, pals: *palustris*, poly: *polyglotta*, rama: *rama*, rufe: *rufescens*, scir: *scirpaceus*, scho: *schoenobaenus*, tait: *taiti*.

Table 1
Comparison of different methods for calculating genetic distances. Comparison between pairwise and complete (bold) deletion. Uncorrected *p*-distances below diagonal and TrN + Γ corrected distances above diagonal.

		1	2	3	4	5	6	7	8	9	10	11	12
1	*A. bistrigiceps* AJ004254		0.034/ **0.020**	0.130/ **0.123**	0.128/ **0.120**	0.129/ **0.099**	0.115/ **0.096**	0.122/ **0.098**	0.136/ **0.110**	0.135/ **0.110**	0.133/ **0.107**	0.131/ **0.114**	0.126/ **0.106**
2	*A. bistrigiceps* AJ004258	0.031/ **0.019**		0.136/ **0.113**	0.134/ **0.110**	0.130/ **0.089**	0.116/ **0.087**	0.123/ **0.089**	0.136/ **0.090**	0.136/ **0.089**	0.134/ **0.087**	0.131/ **0.093**	0.126/ **0.090**
3	*A. paludicola* AJ004290	0.090/ **0.088**	0.091/ **0.082**		0.001/ **0.001**	0.124/ **0.101**	0.105/ **0.094**	0.119/ **0.101**	0.099/ **0.093**	0.103/ **0.098**	0.101/ **0.095**	0.099/ **0.099**	0.099/ **0.096**
4	*A. paludicola* AJ004292	0.088/ **0.086**	0.090/ **0.081**	0.001/ **0.001**		0.121/ **0.099**	0.103/ **0.092**	0.117/ **0.099**	0.097/ **0.090**	0.101/ **0.095**	0.099/ **0.093**	0.097/ **0.096**	0.097/ **0.094**
5	*A. melanopogon mimicus* AJ004275	0.088/ **0.074**	0.086/ **0.068**	0.086/ **0.077**	0.085/ **0.075**		0.028/ **0.018**	0.026/ **0.018**	0.121/ **0.089**	0.126/ **0.093**	0.124/ **0.091**	0.117/ **0.088**	0.117/ **0.086**
6	*A. melanopogon mimicus* AJ004279	0.081/ **0.072**	0.080/ **0.067**	0.077/ **0.072**	0.076/ **0.071**	0.026/ **0.018**		0.009/ **0.005**	0.113/ **0.091**	0.117/ **0.096**	0.115/ **0.093**	0.109/ **0.090**	0.109/ **0.088**
7	*A. melanopogon melanopogon* AJ004282	0.083/ **0.072**	0.082/ **0.067**	0.083/ **0.075**	0.082/ **0.074**	0.024/ **0.018**	0.009/ **0.006**		0.122/ **0.092**	0.127/ **0.097**	0.125/ **0.095**	0.118/ **0.092**	0.118/ **0.089**
8	*A. schoenobaenus* AJ004241	0.088/ **0.078**	0.086/ **0.067**	0.071/ **0.070**	0.070/ **0.068**	0.081/ **0.067**	0.077/ **0.068**	0.081/ **0.068**		0.002/ **0.003**	0.001/ **0.001**	0.006/ **0.003**	0.008/ **0.004**
9	*A. schoenobaenus* AJ004240	0.088/ **0.078**	0.086/ **0.067**	0.074/ **0.072**	0.072/ **0.071**	0.083/ **0.070**	0.080/ **0.071**	0.083/ **0.071**	0.002/ **0.003**		0.001/ **0.001**	0.008/ **0.005**	0.008/ **0.004**
10	*A. schoenobaenus* AJ004239	0.086/ **0.077**	0.085/ **0.065**	0.072/ **0.071**	0.071/ **0.070**	0.082/ **0.068**	0.079/ **0.070**	0.082/ **0.070**	0.001/ **0.001**	0.001/ **0.001**		0.007/ **0.004**	0.007/ **0.003**
11	*A. schoenobaenus* JN574448	0.085/ **0.079**	0.084/ **0.068**	0.071/ **0.072**	0.070/ **0.071**	0.080/ **0.067**	0.076/ **0.068**	0.080/ **0.068**	0.006/ **0.003**	0.008/ **0.006**	0.007/ **0.004**		0.004/ **0.004**
12	*A. schoenobaenus* AJ004244	0.083/ **0.075**	0.082/ **0.064**	0.071/ **0.071**	0.070/ **0.070**	0.080/ **0.065**	0.076/ **0.067**	0.080// **0.067**	0.008/ **0.004**	0.008/ **0.004**	0.007/ **0.003**	0.005/ **0.004**	

Table 2
Comparison of different parts of cytochrome *b*. Comparison of genetic distances of cytochrome *b* using first half (below diagonal) and second half (above diagonal). Different correction methods are given in the following way: uncorrected *p*-distances/TrN + Γ corrected distances (bold). All genetic distances were obtained using the complete deletion alignment of 718 nucleotides.

		1	2	3	4	5	6	7	8	9	10	11	12
1	*A. bistrigiceps* AJ004254		0.022/ **0.023**	0.095/ **0.144**	0.092/ **0.138**	0.078/ **0.114**	0.075/ **0.107**	0.072/ **0.102**	0.078/ **0.113**	0.075/ **0.108**	0.075/ **0.108**	0.081/ **0.121**	0.072/ **0.104**
2	*A. bistrigiceps* AJ004258	0.017/ **0.016**		0.089/ **0.134**	0.086/ **0.128**	0.078/ **0.115**	0.075/ **0.109**	0.072/ **0.103**	0.072/ **0.104**	0.070/ **0.098**	0.070/ **0.098**	0.075/ **0.111**	0.067/ **0.095**
3	*A. paludicola* AJ004290	0.081/ **0.105**	0.075/ **0.095**		0.003/ **0.003**	0.086/ **0.127**	0.078/ **0.111**	0.075/ **0.105**	0.059/ **0.075**	0.061/ **0.080**	0.061/ **0.080**	0.064/ **0.086**	0.061/ **0.082**
4	*A. paludicola* AJ004292	0.081/ **0.105**	0.075/ **0.095**	0.000/ **0.000**		0.084/ **0.122**	0.075/ **0.105**	0.072/ **0.100**	0.056/ **0.071**	0.059/ **0.075**	0.059/ **0.075**	0.061/ **0.081**	0.059/ **0.077**
5	*A. melanopogon mimicus* AJ004275	0.069/ **0.086**	0.058/ **0.068**	0.067/ **0.080**	0.067/ **0.080**		0.025/ **0.027**	0.022/ **0.024**	0.070/ **0.098**	0.072/ **0.103**	0.072/ **0.103**	0.070/ **0.096**	0.067/ **0.091**
6	*A. melanopogon mimicus* AJ004279	0.069/ **0.086**	0.058/ **0.069**	0.067/ **0.080**	0.067/ **0.080**	0.011/ **0.010**		0.003/ **0.003**	0.061/ **0.082**	0.064/ **0.087**	0.064/ **0.087**	0.061/ **0.080**	0.059/ **0.076**
7	*A. melanopogon melanopogon* AJ004282	0.072/ **0.094**	0.061/ **0.076**	0.075/ **0.097**	0.075/ **0.097**	0.014/ **0.014**	0.008/ **0.008**		0.059/ **0.077**	0.061/ **0.082**	0.061/ **0.082**	0.059/ **0.075**	0.056/ **0.071**
8	*A. schoenobaenus* AJ004241	0.078/ **0.107**	0.061/ **0.077**	0.081/ **0.111**	0.081/ **0.111**	0.064/ **0.081**	0.075/ **0.100**	0.078/ **0.109**		0.003/ **0.003**	0.003/ **0.003**	0.006/ **0.006**	0.008/ **0.008**
9	*A. schoenobaenus* AJ004240	0.081/ **0.112**	0.064/ **0.081**	0.083/ **0.116**	0.083/ **0.116**	0.067/ **0.085**	0.078/ **0.104**	0.081/ **0.114**	0.003/ **0.003**		0.000/ **0.000**	0.008/ **0.008**	0.006/ **0.006**
10	*A. schoenobaenus* AJ004239	0.078/ **0.107**	0.061/ **0.077**	0.081/ **0.111**	0.081/ **0.111**	0.064/ **0.081**	0.075/ **0.100**	0.078/ **0.109**	0.000/ **0.000**	0.003/ **0.003**		0.008/ **0.008**	0.006/ **0.006**
11	*A. schoenobaenus* JN574448	0.078/ **0.107**	0.061/ **0.077**	0.081/ **0.111**	0.081/ **0.111**	0.064/ **0.081**	0.075/ **0.100**	0.078/ **0.109**	0.000/ **0.000**	0.003/ **0.003**	0.000/ **0.000**		0.008/ **0.008**
12	*A. schoenobaenus* AJ004244	0.078/ **0.107**	0.061/ **0.077**	0.081/ **0.111**	0.081/ **0.111**	0.064/ **0.081**	0.075/ **0.100**	0.078/ **0.109**	0.000/ **0.000**	0.003/ **0.003**	0.000/ **0.000**	0.000/ **0.000**	

4. Discussion

4.1. Comparison of sequences of unequal lengths

Our results imply that sequences of different lengths are not directly comparable, even if they are from the same locus. This concerns homologous sequences of effectively unequal lengths due to uncertain nucleotides when "pairwise deletion" is used, as well as partially or non-overlapping sequences of the same locus. This is not surprising, since substitution rates are known to vary across a locus, such as the cyt *b* gene (Griffiths, 1997; Klicka et al., 2001). As shown in the present study, conclusions drawn from comparisons of different parts of cyt *b* can lead to contradictory interpretations of taxonomic rank, if a genetic threshold is applied, as e.g. the 2% proposed by Ratnasingham and Hebert (2007). For example, examination of the first half of cyt *b* would support species status of *Acrocephalus mendanae aquilonis* and *Acrocephalus taiti* (0.059), whereas the second part of cyt *b* (0.008) would instead suggest that the divergence is within the range of intraspecific divergence proposed by BOLD (Ratnasingham and Hebert, 2007). In contrast, in *A. bistrigiceps*, analysis of the second half of cyt *b* would suggest much more intrataxon variation than analysis of

the first part. These results are especially important in barcoding, where genetic distances have been proposed to delimit species (Ratnasingham and Hebert, 2007).

Hebert et al. (2003b) also compared both halves of CO1, and found that the mean sequence divergence in the 5′ part was 97.7% of the 3′ end, with a standard deviation of 6.2%. Thus, the divergence between both halves of CO1 is greater than the threshold of 2% for species delimitation according to BOLD (Ratnasingham and Hebert, 2007). Nevertheless, Hebert et al. (2003b) concluded that "because of this congruence, the measures of sequence divergence for other species pairs are analysed without reference to their source region in the gene". However, our analyses of first vs. second halves of cyt *b* demonstrated that generalizations can be misleading. Unfortunately, the lengths of compared sequences are often not stated. Instead, they are frequently given as minimum values (e.g. larger than a given number of base pairs; Aliabadian et al., 2009; Nijman and Aliabadian, 2010), as a range of sequence lengths (Alexander et al., 2009; Kerr et al., 2009b; Virgilio et al., 2010; Wiemers and Fiedler, 2007), or as mean values of sequence lengths (Kerr et al., 2007, 2009a). Alternatively, other sequence lengths than the proposed 648 base pairs of CO1 for barcoding are used (Langhoff et al., 2009; Trewick, 2008; citations within Waugh, 2007; Whitworth et al., 2007). Roe and Sperling (2007) studied CO1 and CO2, and found substantial differences between different parts of the same gene and also between these genes. They suggested to maximize sequence length for calculating genetic divergences to reduce the stochastic variation that occurs when short sequences are used.

Genetic distances were generally higher in "pairwise deletion" than in "complete deletion", reflecting the smaller amount of data analyzed by the latter method. As a result of the differences between these methods, distances calculated by different methods are not directly comparable. Moreover, for any group of taxa with sequences of varying completeness, distances calculated by "pairwise deletion" are not directly comparable among these taxa. For example, given three equally distant species A, B and C and shortening the sequence of the latter, "pairwise deletion" will almost always result in relatively greater differences between A and B than between any of these and C. In contrast, "complete deletion" will yield equal distances. Unfortunately, not all programs offer a choice between these two methods (Treefinder (Jobb et al., 2004; Jobb, 2008): pairwise deletion; PAUP* (Swofford, 2003): mean differences and pairwise deletion, MEGA (Tamura et al., 2007): choice between pairwise and complete deletion).

We conducted a survey of cyt *b* sequences of birds in GenBank on 25.05.2010, which revealed that the majority is incomplete: only ~16% of a total of ~19,300 sequences may be considered complete (≥1140 bp). Over one third (35%) are at best 800 bp long, 22% range from 801 to 1000 bp and 27% are 1001–1039 bp long. In other words, most of these are not directly comparable. Recalculating genetic distances with sequences trimmed to equal length, as done by Barrett and Hebert (2005) and Klicka et al. (2007), is not common practice.

In conclusion, our results imply that genetic distances are not comparable if different sequence lengths and/or different parts of the same locus are involved. This is true also for comparisons of sequences that include uncertain nucleotides, if "pairwise deletion" is used. Comparing genetic distances from different loci or distances obtained by different methods (e.g. sequence data vs. allozyme data) is even more problematic, since different loci might have strongly differing substitution rates (e.g. Palma et al., 2010; Roe and Sperling, 2007).

4.2. Uncorrected vs. corrected distances

The paradox of choosing the appropriate distance measure.

Uncorrected distances inevitably underestimate genetic distances if multiple hits have occurred. For this reason, distance corrections have been introduced (Swofford et al., 1996). For phylogenetic analyses it has been emphasized that selecting the optimal substitution model is pertinent (Johnson and Omland, 2003; Lemmon and Moriarty, 2004; Posada and Buckley, 2004; Posada, 2008). It would thus seem advisable to exercise the same care when selecting the model for calculation of genetic distances, which, however, is rarely done (exceptions e.g. Alström et al., 2007; Haase et al., 2007; Olsson et al., 2004; Schenk and Hufford, 2010; Svensson et al., 2008; Trewick, 2000; Waters et al., 2006). However, genetic distances are frequently calculated using a simpler than the best-fit model for the particular dataset (Arshad et al., 2009; Banks et al., 2006; Burns and Naoki, 2004; Feinstein et al., 2008; Gill et al., 2005; Klicka et al., 2005; Palma et al., 2010; Woog et al., 2008; Zhang et al., 2007), and the Kimura-two-parameter (Kimura, 1980) model is especially common in barcoding studies (Barrett and Hebert, 2005; Clare et al., 2007; Efe et al., 2009; Hebert et al., 2003a; Hubert et al., 2008; Linares et al., 2009). Furthermore, uncorrected ("p") values are often used (Alström et al., 2008a; Banks et al., 2006; Bates et al., 1999; Benzoni et al., 2010; Dietzen et al., 2008; Langhoff et al., 2009; Loader et al., 2010; Penhallurick and Wink, 2004; Qu et al., 2006). The problem of distance correction seems to be more severe for older divergences than for recent ones (e.g. Lemmon and Moriarty, 2004; Posada and Crandall, 2001; Sullivan and Swofford, 1997; Yang et al., 1994), as has also been shown here. Thus, for species delimitation and barcoding, it has been suggested that a model comprising only few parameters such as the Kimura 2 parameter model (K2P, Kimura, 1980) would be appropriate (Hebert et al., 2003a; Nei and Kumar, 2000). However, identical uncorrected differences between pairs of sequences might translate into substantially different values using a correction method (Fig. 2). The amount and direction (sign) of the difference may depend on the substitution model. Therefore, arbitrarily choosing a substitution model leaves us with the uncertainty whether distances have been corrected accurately. However, this issue is crucial when a threshold decides if two organisms belong to the same species or not. Doubts about the appropriateness of indiscriminately using the K2P model in barcoding have been raised earlier, but often this distance correction is still used for reasons of comparability (e.g. Casiraghi et al., 2010; Derycke et al., 2010; Wiemers and Fiedler, 2007). However, the only common basis in such comparisons is the uncertainty. Likewise, indiscriminately comparing sequence divergences obtained using different arbitrary models, as is often done (e.g. Arshad et al., 2009; Banks et al., 2006; Burns and Naoki, 2004; Feinstein et al., 2008; Gill et al., 2005; Helbig et al., 1995; Klicka et al., 2005; Woog et al., 2008; Zhang et al., 2007), is almost certainly misleading and should also be avoided.

We recommend that genetic distances used in species discrimination be based on best-fit substitution models, but using a universal cut-off value is not advisable in any case. Only accurately corrected distances can be compared across data sets, even if the correcting models differ for these data sets. However, ignoring the fact that the optimal model will often not be the true model underlying sequence evolution (Buckley and Cunningham, 2002; Posada and Buckley, 2004), the practice of model fitting reveals another problem. Ideally, we want to know the number of mutations including multiple hits that separate two sequences. However, estimating the best-fit model for each pairwise comparison in an alignment is not possible. As the choice of model depends on the taxonomic composition of an alignment, the genetic distance between a particular pair of sequences might vary depending on the context. In other words, if different model parameters or different models are estimated for different alignments that include a particular pair of sequences, the distances between them will not

be the same. Thus, corrected genetic distances are unique to the study in which they are calculated as they are conditional on the particular dataset and model selected for that dataset. This raises the question of choosing the appropriate taxonomic context for species discrimination based on genetic distances including barcoding. However, any selection of taxa incurs subjectivity that will inevitably influence distance estimates.

In this paradoxical situation, the inevitable error of distance estimation is probably minimized by the use of best-fit models, a notion that can and should be tested in simulations. However, considering the numerous pitfalls for distance estimation outlined here as well as in the previous sections and by foregoing studies (Palma et al., 2010; Roe and Sperling, 2007; Schenk and Hufford, 2010) it is advisable not to rely solely on the comparison of genetic distances against a certain threshold in species discrimination, regardless of the correction method. The number of scientists aware of these problems and their calls for caution and integrative approaches in taxonomy is increasing (e.g. Alström et al., 2008b; Capa et al., 2010; Ekrem et al., 2010; Elias et al., 2007; Gomez et al., 2007; Haase et al., 2007; Ratnasingham and Hebert, 2007).

5. Taxonomic comments

We did not find a "barcoding gap" between intra- and interspecific divergences, as proposed by Aliabadian et al. (2009) for cyt *b*. However, as discussed below, this might, at least in part, be explained by inconsistent taxonomic treatments. Moreover, taxonomic ranking based on a threshold value (such as the 2% proposed by BOLD [Barcoding of Life Data System, Ratnasingham and Hebert, 2007] for CO1) was affected by the choice of correction method.

Some taxa were unexpectedly divergent, whereas others were surprisingly similar. In particular some Pacific taxa that are treated as separate species, e.g. *Acrocephalus mendanae* and *Acrocephalus aequinoctialis*, differed very slightly (0.003–0.011, TrN + Γ), whereas others that are considered to be subspecies of the same species, e.g. *Acrocephalus kerearako kerearako* and *Acrocephalus kerearako kaoko*, were much more divergent (0.017, TrN + Γ). Conversely, three Asian mainland species considered to be monotypic, *Acrocephalus bistrigiceps*, *Acrocephalus dumetorum* and the recently rediscovered *Acrocephalus orinus* (Round et al., 2007; Svensson et al., 2008), displayed unexpectedly high intraspecific genetic divergences.

6. Conclusions

Genetic distances should be carefully calculated and compared. We suggest that (1) different loci are incomparable; (2) only perfectly homologous regions (same length, same part of locus) should be compared; (3) in the case of sequences with some uncertain nucleotides, only distances calculated by the method of "complete deletion" are fully comparable. (4) Only distances based on the optimal substitution model should be used. (5) Even within the same locus, corrected genetic distances are unique to the study in which they are calculated, as they are conditional on the particular dataset and model selected for that dataset.

Acknowledgments

We are grateful to H. Shirihai, L. Svensson, P.D. Round for providing samples. This study was partially financed by a European Union Synthesys grant (SE-TAF-2992 to S.F.) and by The Swedish Research Council (Grant No. 621-2006-3194 to U.O.). We are also thankful for comments from anonymous reviewers on this manuscript.

Appendix A. Taxon list

Taxonlist in alphabetical order. Accession numbers, locality and references for sequences downloaded from Genbank are given.

Taxon	Accession No.	Locality	References
Acrocephalus aequinoctialis aequinoctialis	EF156277	Kiribati Is., Kirimati (Christmas)	Cibois et al. (2007)
Acrocephalus aequinoctialis aequinoctialis	EF156278	Kiribati Is., Kirimati (Christmas)	Cibois et al. (2007)
Acrocephalus agricola	JN574449	Bulgaria	This study
Acrocephalus agricola agricola	AJ004246	Kazakhstan	Leisler et al. (1997)
Acrocephalus agricola agricola	AJ004775	China	Helbig and Seibold (1999)
Acrocephalus arundinaceus zarudnyi	AJ004252	Kazakhstan	Leisler et al. (1997)
Acrocephalus arundinaceus zarudnyi	JN574447	China, Xinjiang	This study
Acrocephalus atyphus ssp.	EF156281	Tuamotu Is., Takapoto	Cibois et al. (2007)
Acrocephalus australis ssp.	AJ004305	Australia	Leisler et al. (1997)
Acrocephalus baeticatus guiersi/cinnamomeus	AJ004234	Senegal	Leisler et al. (1997)
Acrocephalus baeticatus hallae	FJ883024	South Africa	Fregin et al. (2009)
Acrocephalus bistrigiceps	AJ004254	Thailand	Leisler et al. (1997)
Acrocephalus bistrigiceps	AJ004258	Ussuria	Leisler et al. (1997)
Acrocephalus brevipennis	FJ883026	Cape Verde	Fregin et al. (2009)
Acrocephalus brevipennis	AJ004259	Cape Verde	Leisler et al. (1997)
Acrocephalus caffer	EF156308	Society Is., Tahiti,	Cibois et al. (2007)
Acrocephalus concinens concinens	FJ883027	Thailand	Fregin et al. (2009)
Acrocephalus concinens concinens	AJ004260	Thailand	Leisler et al. (1997)
Acrocephalus concinens concinens	AJ004262	Thailand	Leisler et al. (1997)
Acrocephalus dumetorum	JN574458	India, Harike	This study
Acrocephalus dumetorum	JN574459	Nepal	This study
Acrocephalus dumetorum	AJ004263	Kazakhstan	Leisler et al. (1997)
Acrocephalus dumetorum	AJ004264	Finland	Leisler et al. (1997)

Appendix A (*continued*)

Taxon	Accession No.	Locality	References
Acrocephalus familiaris kingi	EU119965	Nihoa Island	Fleischer et al. (2007)
Acrocephalus gracilirostris gracilirostris	AJ004267	South Africa	Leisler et al. (1997)
Acrocephalus gracilirostris ssp.	AJ004266	Kenya	Leisler et al. (1997)
Acrocephalus gracilirostris ssp.	AJ004270	Kenya	Leisler et al. (1997)
Acrocephalus griseldis	AJ004272	Kenya	Leisler et al. (1997)
Acrocephalus kerearako kaoko	AJ004272	Cook Is., Mitiaro	Cibois et al. (2007)
Acrocephalus kerearako kerearako	EF156292	Cook Is., Mangaia	Cibois et al. (2007)
Acrocephalus longirostris	EU303308	Móorea	Cibois et al. (2008)
Acrocephalus melanopogon melanopogon	AJ004282	Austria	Leisler et al. (1997)
Acrocephalus melanopogon mimicus	AJ004275	Kazakhstan	Leisler et al. (1997)
Acrocephalus melanopogon mimicus	AJ004279	Kazakhstan	Leisler et al. (1997)
Acrocephalus mendanae aquilonis	EF156279	Marquesas Is., Eiao	Cibois et al. (2007)
Acrocephalus mendanae aquilonis	EF156280	Marquesas Is., Eiao	Cibois et al. (2007)
Acrocephalus mendanae consobrinus	EF156282	Marquesas Is., Mohotani	Cibois et al. (2007)
Acrocephalus mendanae dido	EF156311	Marquesas Is., Ua Pou	Cibois et al. (2007)
Acrocephalus mendanae fatuhivae	EF156283	Marquesas Is., Fatu Iva	Cibois et al. (2007)
Acrocephalus mendanae idae	EF156290	Marquesas Is., Ua Huka	Cibois et al. (2007)
Acrocephalus mendanae idae	EF156313	Marquesas Is., Ua Huka	Cibois et al. (2007)
Acrocephalus mendanae mendanae	EF156288	Marquesas Is., Hiva Oa	Cibois et al. (2007)
Acrocephalus mendanae mendanae	EF156289	Marquesas Is., Hiva Oa	Cibois et al. (2007)
Acrocephalus mendanae mendanae	EF156293	Marquesas Is., Tahuata	Cibois et al. (2007)
Acrocephalus mendanae percernis	EF156298	Marquesas Is., Nuku Hiva	Cibois et al. (2007)
Acrocephalus mendanae percernis	EF156300	Marquesas Is., Nuku Hiva	Cibois et al. (2007)
Acrocephalus mendanae postremus	EF156303	Marquesas Is., Hatuta'a	Cibois et al. (2007)
Acrocephalus mendanae postremus	EF156304	Marquesas Is., Hatuta'a	Cibois et al. (2007)
Acrocephalus musae garretti	EU303306	Huahine	Cibois et al. (2008)
Acrocephalus musae musae	EU303310	Raiatea	Cibois et al. (2008)
Acrocephalus newtoni	AJ004283	Madagascar	Leisler et al. (1997)
Acrocephalus orientalis	JN574459	China	This study
Acrocephalus orientalis	AB159181	Japan, Osaka	Nishiumi and Kim (2004)
Acrocephalus orientalis	AB159183	Japan, Hiroshima	Nishiumi and Kim (2004)
Acrocephalus orientalis	AB159185	Korea, Gyeonggi-do	Nishiumi and Kim (2004)
Acrocephalus orientalis	AB159186	Korea, Gyeonggi-do	Nishiumi and Kim (2004)
Acrocephalus orientalis	AJ004286	Philippines	Leisler et al. (1997)
Acrocephalus orinus	JN574457	Pakistan, Gilgit	This study
Acrocephalus orinus	JN574451	Afghanistan	This study
Acrocephalus orinus	JN574455	Kazakhstan	This study
Acrocephalus orinus	JN574452	Burma	This study
Acrocephalus orinus	JN574456	Burma	This study
Acrocephalus orinus	JN574453	Afghanistan	This study
Acrocephalus orinus	JN574454	Afghanistan	This study
Acrocephalus orinus	DQ681065	Thailand	Round et al. (2007)
Acrocephalus paludicola	AJ004290	Poland	Leisler et al. (1997)
Acrocephalus paludicola	AJ004292	Poland	Leisler et al. (1997)
Acrocephalus palustris	AJ004294	Oman	Leisler et al. (1997)
Acrocephalus rimitarae	EF156305	Austral Is., Rimatara	Cibois et al. (2007)
Acrocephalus rimitarae	EF156306	Austral Is., Rimatara	Cibois et al. (2007)
Acrocephalus rufescens ansorgei	AJ004295	Kenya	Leisler et al. (1997)
Acrocephalus rufescens ansorgei	AJ004296	Kenya	Leisler et al. (1997)
Acrocephalus rufescens senegalensis	FJ883037	Senegal	Fregin et al. (2009)
Acrocephalus schoenobaenus	AJ004239	Germany	Leisler et al. (1997)
Acrocephalus schoenobaenus	AJ004240	Germany	Leisler et al. (1997)
Acrocephalus schoenobaenus	AJ004241	Oman	Leisler et al. (1997)
Acrocephalus schoenobaenus	AJ004244	unknown	Leisler et al. (1997)
Acrocephalus schoenobaenus	JN574448	China, Xinjiang	This study
Acrocephalus scirpaceus avicenniae	AJ004237	Saudi Arabia	Leisler et al. (1997)
Acrocephalus scirpaceus fuscus	AJ004297	Kazakhstan	Leisler et al. (1997)
Acrocephalus scirpaceus fuscus	AJ004298	Kenya	Leisler et al. (1997)
Acrocephalus scirpaceus scirpaceus	JN574450	Bulgaria	This study

(*continued on next page*)

Appendix A (*continued*)

Taxon	Accession No.	Locality	References
Acrocephalus scirpaceus scirpaceus	AJ004301	Germany	Leisler et al. (1997)
Acrocephalus scirpaceus scirpaceus	AJ004304	unknown	Leisler et al. (1997)
Acrocephalus sechellensis	AJ004284	Seychelles	Leisler et al. (1997)
Acrocephalus stentoreus amyae	JN574443	Burma	This study
Acrocephalus stentoreus brunnescens	AJ004307	Oman	Leisler et al. (1997)
Acrocephalus stentoreus brunnescens	JN574444	India, Harike	This study
Acrocephalus stentoreus harterti	JN574445	Philippines	This study
Acrocephalus stentoreus harterti	JN574446	Philippines	This study
Acrocephalus stentoreus levantinus	JN574442	Israel, Dead Sea	This study
Acrocephalus taiti	AJ004308	Henderson Isl.	Leisler et al. (1997)
Acrocephalus taiti	AJ004309	Henderson Isl.	Leisler et al. (1997)
Calamonastides gracilirostris	FJ883043	Kenya	Fregin et al. (2009)
Hippolais icterina	AJ004316	unknown	Leisler et al. (1997)
Hippolais icterina	DQ008479	Sweden	Alström et al. (2006)
Hippolais languida	JN574472	Kenya	This study
Hippolais olivetorum	JN574473	Kenya	This study
Hippolais olivetorum	JN574474	Bulgaria	This study
Hippolais polyglotta	AF094619	France	Cibois et al. (1999)
Iduna aedon ssp.	AF094623	Thailand, Umphang,	Cibois et al. (1999)
Iduna caligata	AJ004315	unknown	Leisler et al. (1997)
Iduna caligata	JN574461	Finland	This study
Iduna caligata	JN574460	Kazakhstan	This study
Iduna natalensis batsi	AF094620	Cameroon	Cibois et al. (1999)
Iduna natalensis natalensis	DQ008523	Malawi	Alström et al. (2006)
Iduna opaca	AJ004317	Unknown	Leisler et al. (1997)
Iduna opaca	JN574466	Senegal	This study
Iduna opaca	JN574467	Marocco	This study
Iduna pallida elaeica	JN574468	Kazakhstan	This study
Iduna pallida elaeica	JN574469	Kazakhstan	This study
Iduna pallida elaeica	JN574470	Kazakhstan	This study
Iduna rama	AJ004792	Turkmenistan	Helbig and Seibold (1999)
Iduna rama	JN574462	Kazakhstan	This study
Iduna rama	JN574463	Kazakhstan	This study
Iduna rama	JN574464	Kazakhstan	This study
Iduna rama	JN574465	Kazakhstan	This study
Iduna similis	JN574471	Tanzania	This study
Iduna similis	FJ899738	Kenya	Fregin et al. (2009)

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.10.003.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automatic Control 19, 716–723.

Alexander, L.C., Delion, M., Hawthorne, D.J., Lamp, W.O., Funk, D.H., 2009. Mitochondrial lineages and DNA barcoding of closely related species in the mayfly genus *Ephemerella* (Ephemeroptera: Ephemerellidae). J. N. Am. Benthol. Soc. 28 (3), 584–595.

Aliabadian, M., Kaboli, M., Nijman, V., Vences, M., 2009. Molecular identification of birds: performance of distance-based DNA barcoding in three genes to delimit parapatric species. PLoS ONE 4 (1), e4119. doi:10.1371/journal.pone.0004119.

Alström, P., Ericson, P.G.P., Olsson, U., Sundberg, P., 2006. Phylogeny and classification of the avian superfamily Sylvioidea. Mol. Phylogenet. Evol. 38, 381–397.

Alström, P., Olsson, U., Rasmussen, P.C., Yao, C.-T., Ericson, P.G.P., Sundberg, P., 2007. Morphological, vocal and genetic divergence in the *Cettia acanthizoides* complex (Aves: Cettiidae). Zool. J. Linn. Soc. 149, 437–452.

Alström, P., Olsson, U., Lei, F., Wang, H., Gao, W., Sundberg, P., 2008a. Phylogeny and classification of the Old World Emberizini (Aves, Passeriformes). Mol. Phylogenet. Evol. 47, 960–973.

Alström, P., Rasmussen, P.C., Olsson, U., Sundberg, P., 2008b. Species delimitation based on multiple criteria: the Spotted Bush Warbler *Bradypterus thoracicus* complex (Aves, Megaluridae). Zool. J. Linn. Soc. 154, 291–307.

Arbogast, B.S., Edwards, S.V., Wakeley, J., Beerli, P., Slowinski, J.B., 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annu. Rev. Ecol. Syst. 33, 707–740.

Arshad, M., Gonzalez, J., El-Ayed, A.A., Osborne, T., Wink, M., 2009. Phylogeny and phylogeography of critically endangered *Gyps* species based on nuclear and mitochondrial markers. J. Ornithol. 150, 419–430.

Bairlein, F., Alström, P., Aymi, R., Clement, P., Dyrcz, A., Gargallo, G., Hawkins, F., Madge, S., Pearson, D., Svensson, L., 2006. Family Sylviidae (Old World Warblers). In: del Hoyo, J., Elliott, A., Christie, D.A. (Eds.), Handbook of the Birds of the World, vol. 11. Old World Flycatchers to Old World Warblers, Lynx Edicions, Barcelona, pp. 492–709.

Banks, J., van Buren, A., Cherel, Y., Whitfield, J.B., 2006. Genetic evidence for three species of rockhopper penguins, *Eudyptes chrysocome*. Polar Biol. 30, 61–67.

Baker, J.M., Lopez-Medrano, E., Navarro-Siguenza, A.G., Rojas-Soto, O.R., Omland, K.E., 2003. Recent speciation in the orchard oriole group: divergence of *Icterinus spurius spurius* and *Icterus spurius fuertesi*. Auk 120 (3), 848–859.

Barrett, R.D.H., Hebert, P.D.N., 2005. Identifying spiders through DNA barcodes. Can. J. Zool. 83, 481–491.

Bates, J.M., Hackett, S.J., Goerck, J.M., 1999. High levels of mitochondrial DNA differentiation in two lineages of antbirds (*Dryomophila* and *Hypocnemis*). Auk 116 (4), 1093–1106.

Benzoni, F., Stefani, F., Pichon, M., Galli, P., 2010. The name game: morpho-molecular species boundaries in the genus *Psammocora* (Cnidaria, Scleratinia). Zool. J. Linn. Soc. 160, 421–456.

Besansky, N.J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, O., Fontenille, D., Toure, Y., Sagnon, N.F., 2003. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. PNAS 100 (19), 10818–10823.

Bradley, R.D., Baker, R.J., 2001. A test of the genetic species concept: cytochrome-b sequences and mammals. J. Mammal. 82 (4), 960–973.

Broughton, R.E., Reneau, P.C., 2006. Spatial covariation of mutation and nonsynonymous substitution rates in vertebrae mitochondrial genomes. Mol. Biol. Evol. 23 (8), 1516–1524.
Buckley, T.R., Cunningham, C.W., 2002. The effect of nucleotide substitution model assumption on estimates of nonparametric bootstrap support. Mol. Biol. Evol. 19 (4), 394–405.
Burbrink, F.T., Lawson, R., Slowinski, J.B., 2000. Mitochondrial DNA phylogeography of the polytypic North American Rat Snake (*Elaphe obsoleta*): a critique of the subspecies concept. Evolution 54 (6), 2107–2118.
Burns, K.J., Naoki, K., 2004. Molecular phylogenetics and biogeography of Neotropical tanagers in the genus *Tangara*. Mol. Phylogenet. Evol. 32, 838–854.
Cagnon, C., Lauga, B., Hémery, G., Mouchès, C., 2004. Phylogeographic differentiation of storm petrels (*Hydrobates pelagicus*) based on cytochrome b mitochondrial DNA variation. Mar. Biol. 145, 1257–1264.
Capa, M., Bybee, D.R., Bybee, S.M., 2010. Establishing species and species boundaries in *Sabellastarte* Krøyer, 1856 (Annelidae: Sabellidae): an integrative approach. Org. Divers. Evol. 10, 351–371.
Casiraghi, M., Labra, M., Ferri, E., Galimberti, A., De Mattia, F., 2010. DNA barcoding: a six- question tour to improve users' awareness about the method. Brief. Bioinform. 4 (2), 440–453.
Cibois, A., Pasquet, E., Schulenberg, T.S., 1999. Molecular systematics of the Malagasy babblers (Passeriformes: Timaliidae) and Warblers (Passeriformes: Sylviidae), based on cytochrome *b* and 16S rRNA sequences. Mol. Phylogenet. Evol. 13 (3), 581–595.
Cibois, A., Thibault, J.C., Pasquet, E., 2007. Uniform phenotype conceals double colonization by reed-warblers of a remote Pacific archipelago. J. Biogeogr. 34 (7), 1155–1166.
Cibois, A., Thibault, J.C., Pasquet, E., 2008. Systematic of the extinct reed warblers *Acrocephalus* of the Society Islands of Eastern Polynesia. Ibis 150, 365–376.
Cicero, C., Johnson, N.K., 1995. Speciation in Sapsuckers (*Spyrapicus*): III. Mitochondrial-DNA sequence divergence at the cytochrome-b locus. Auk 112 (3), 547–563.
Clare, E.L., Lim, B.K., Engstrom, M.D., Eger, J.L., Hebert, P.D.N., 2007. DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Mol. Ecol. Notes 7, 184–190.
Cognato, A.I., 2006. Standard percent DNA sequence difference for insects does not predict species boundaries. J. Econ. Entomol. 99 (4), 1037–1045.
Cramp, S., Brooks, D.J., Dunn, E., Gillmor, R., Hall-Craggs, J., Hollom, P.A.D., Nicholson, E.M., Ogilvie, M.A., Roselaar, C.S., Sellar, P.J., Simmons, K.E.L., Snow, D.W., Vincent, D., Voous, K.H., Wallace, D.I.M., Wilson, M.G., 1992. Warblers. In: Brooks, D.J. (Ed.), The Birds of the Western Palearctic, vol. 7. Oxford University Press, New York, pp. 106–309.
Dalebout, M.L., Baker, S.C., Steel, D., Robertson, K.M., Chivers, S.J., Perrin, W.F., Mead, J.G., Brace, R.V., Schofield Jr., T.D., 2007. A divergent mtDNA lineage among Mesoplodon Beaked Whales: molecular evidence for a new species in the tropical Pacific? Mar. Mammal. Sci. 23 (4), 954–966.
Davis, J.I., Nixon, K.C., 1992. Populations, genetic variation, and the delimitation of phylogenetic species. Syst. Biol. 41, 421–435.
Davison, A., Blackie, R.L.E., Scothern, G.P., 2009. DNA barcoding of stylommatophoran land snails: a test of existing sequences. Mol. Ecol. Resour. 9, 1092–1101.
Derycke, S., Vanaverbeke, J., Rigaux, A., Backeljau, T., Moens, T., 2010. Exploring the use of cytochrome oxidase c subunit 1 (CO1) for DNA barcoding of free-living marine nematodes. PloS ONE 5 (10), e13716. doi:10.1371/journal.pone.0013716.
Dietzen, C., Garcia-del-Rey, E., Delgado Castro, G., Wink, M., 2008. Phylogeography of the blue tit (*Parus teneriffae*-group) on the Canary Islands based on mitochondrial DNA sequence data and morphometrics. J. Ornithol. 149, 1–12.
Efe, M.A., Tavares, E.S., Baker, A.J., Bonatto, S.L., 2009. Multigene phylogeny and DNA barcoding indicate that the Sandwich tern complex (*Thalasseus sandvicencis*, Laridae, Sternini) comprises two species. Mol. Phylogenet. Evol. 52, 263–267.
Ekrem, T., Stur, E., Hebert, P.D., 2010. Females do count: documenting chironomidae (Diptera) species diversity using DNA barcoding. Org. Divers. Evol. 10, 397–408.
Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J., Jiggins, C.D., 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc. R. Soc. B. 274, 2881–2889.
Feinstein, J., Yang, X., Li, S.H., 2008. Molecular systematics and historical biogeography of the Black-browed Barbet species complex (*Megalaima oorti*). Ibis 150, 40–49.
Fleischer, R.C., Slikas, B., Beadell, J., Atkins, C., McIntosh, C.E., Conant, C., 2007. Genetic variability and taxonomic status of the Nihoa and Laysan Millerbirds. Condor 109 (4), 954–962.
Fregin, S., Haase, M., Olsson, U., Alström, P., 2009. Multilocus phylogeny of the family Acrocephalidae–The traditional taxonomy overthrown. Mol. Phylogenet. Evol. 52, 866–878.
Galtier, N., Enard, D., Radondy, Y., Bazin, E., Belkhir, K., 2006. Mutation hot spots in mammalian mitochondrial DNA. Genome. Res. 16, 215–222.
Gill, F.B., Slikas, B., Sheldon, F.H., 2005. Phylogeny of titmice (Paridae): II. Species relationships based on sequences of the mitochondrial cytochrome-b gene. Auk 122 (1), 121–143.
Gomez, A., Serra, M., Carvalho, G.R., Lunt, D.H., 2002. Speciation in ancient cryptic species complexes: evidence from the molecular phylogeny of *Brachionus plicatilis* (Rotifera). Evolution 56 (7), 1431–1444.
Gomez, A., Wright, P.J., Lunt, D.H., Cancino, J.M., Carvalho, G.R., Hughes, R.N., 2007. Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan taxon. Proc. R. Soc. B 274, 99–207.
Griffiths, C.S., 1997. Correlations of functional domains and rates of nucleotide substitution in cytochrome-b. Mol. Phylogenet. Evol. 5, 368–382.
Haase, M., 2005. Rapid and convergent evolution of parental care in hydrobiid gastropods from New Zealand. J. Evol. Biol. 18, 1076–1086.
Haase, M., Marshall, B., Hogg, I., 2007. Disentangling causes of disjunction on the South Island of New Zealand: the Alpine fault hypothesis of vicariance revisited. Biol. J. Linn. Soc. 91, 361–374.
Hajibabaei, M., Janzan, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D.N., 2006. DNA barcodes distinguish species of tropical Lepidoptera. PNAS 103 (4), 968–971.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic. Acids. Symp. Ser. 41, 95–98.
Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003a. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B Biol. Sci. 270, 313–321.
Hebert, P.D.N., Ratnasingham, S., DeWaard, J.R., 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. Lond. B Biol. Sci. 270, S596–S599.
Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of birds through DNA barcodes. PloS Biol. 2 (10), e312.
Hebert, P.D.N., deWaard, J.R., Landry, J.-F., 2010. DNA barcodes for 1/1000 of the animal kingdom. Biol. Lett. 6, 359–362.
Helbig, A.J., Seibold, I., Martens, J., Wink, M., 1995. Genetic differentiation and phylogenetic relationships of Bonellís Warbler *Phylloscopus bonelli* and Green Warbler *P. Nitidus*. J. Avian Biol. 26, 139–153.
Helbig, A.J., Martens, J., Seibold, I., Henning, F., Schottler, B., Wink, M., 1996. Phylogeny and species limits in the palaearctic chiffchaff *Phylloscopus collybita* complex: mitochondrial genetic differentiation and bioacoustic evidence. Ibis 138, 650–666.
Helbig, A.J., Seibold, I., 1999. Molecular Phylogeny of Palearctic–African *Acrocephalus* and *Hippolais* Warblers (Aves: Sylviidae). Mol. Phylogenet. Evol. 11, 246–260.
Hickerson, M.J., Gilchrist, M.A., Takebayashi, N., 2003. Calibrating a molecular clock from phylogeographic data: moments and likelihood estimators. Evolution 57, 2216–2225.
Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J., Bernatchez, L., 2008. Identifying canadian freshwater fishes through DNA barcodes. PLoS ONE 3 (6), e2490. doi:10.1371/journal.pone.0002490.
Hung, G.-C., Chilton, N.B., Beveridge, I., Zhu, X.Q., Lichtenfels, J.R., Gasser, R.G., 1999. Molecular evidence for cryptic species within *Cylicostephanus minutus* (Nematoda: Strongylidae). Int. J. Parasitol. 29, 285–291.
Hunt, B., Strugnell, J., Bednarsek, N., Linse, K., Nelson, R.J., Pakhomov, E., Seibel, B., Steinke, D., Würzberg, L., 2010. Poles apart: The ''bipolar'' pteropod species *Limacina helicina* is genetically distinct between the Arctic and Antarctic Oceans. PLoS ONE 5 (3), e9835. doi:10.1371/journal.pone.0009835.
Irwin, D.E., Rubtsov, A.S., Panov, E.N., 2009. Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). Biol. J. Linn. Soc. 98, 422–438.
Jobb, G., von Haeseler, A., Strimmer, K., 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4, 1471–2148.
Jobb, G., 2008. TREEFINDER version of October. Munich, Germany. <www.treefinder.de>.
Johansson, U., Fjeldså, J., Bowie, R., 2008. Phylogenetic relationships within Passerida (Aves: Passeriformes): a review and a new molecular phylogeny based on three nuclear intron markers. Mol. Phylogenet. Evol. 48, 858–876.
Johnsen, A., Rindal, E., Ericson, P.G.P., Zuccon, D., Kerr, K.C.R., Stoeckle, M.Y., Lifjeld, J.T., 2010. DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. J. Ornithol. 151, 565–578.
Johnson, J.B., Omland, K.S., 2003. Model selection in ecology and evolution. Trends Ecol. Evol. 19, 101–108.
Johnson, N.K., Cicero, C., 2004. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. Evolution 58, 1122–1130.
Kergoat, G.J., Le Ru, P.B., Genson, G., Cruaud, C., Couloux, A., Delobel, A., 2011. Phylogenetics, species boundaries and timing of resource tracking in a highly specialized group of seed beetles (Coleoptera: Chrysomelidae: Bruchinae). Mol. Phylogenet. Evol. 59, 746–760.
Kerr, K.C.R., Stoeckle, M.Y., Dove, C.J., Weigt, L.E., Francis, C.M., Hebert, P.D.N., 2007. Comprehensive DNA barcode coverage of North American birds. Mol. Ecol. Notes 7, 535–543.
Kerr, K.C.R., Lijtmaer, D.A., Barreira, A.S., Hebert, P.D.N., Tubaro, P.L., 2009a. Probing evolutionary patterns in neotropical birds through DNA barcodes. PloS ONE 4 (2), e4379. doi:10.1371/journal.pone.0004379.
Kerr, K.C.R., Birks, S.M., Kalyakin, M.V., Red'kin, Y.A., Koblik, E.A., Hebert, P.D.N., 2009b. Filling the gap – COI barcode resolution in eastern Palearctic birds. Front. Zool. 6, 29. doi:10.1186/1742-9994-6-29.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Klicka, J., Zink, R.M., Barlow, J.C., McGillivray, W.B., Doyle, T.J., 1999. Evidence supporting the recent origin and species status of the Timberline Sparrow. Condor 101, 577–588.
Klicka, J., Fry, A.J., Zink, R.M., Thompson, C.W., 2001. A cytochrome-*b* perspective on *Passerina* Bunting relationships. Auk 118 (3), 611–623.
Klicka, J., Voelker, H., Spellman, G.M., 2005. A molecular phylogenetic analysis of the ''true thrushes''. Mol. Phyogenet. Evol. 34, 486–500.

Klicka, J., Burns, K., Spellman, G.M., 2007. Defining a monophyletic Cardinalini: a molecular perspective. Mol. Phylogenet. Evol. 45, 1010–1032.
Langhoff, P., Authier, A., Buckley, T.R., Dugdale, J.S., Rodrigos, A., Newcomb, R.D., 2009. DNA barcoding of the endemic New Zealand leafroller moth genera, *Ctenopseustis* and *Planotortrix*. Mol. Ecol. Resour. 9, 691–698.
Leisler, B., Heidrich, P., Schulze-Hagen, K., Wink, M., 1997. Taxonomy and phylogeny of reed warblers (genus *Acrocephalus*) based on mtDNA sequences and morphology. J. Ornithol. 138, 469–496.
Lemmon, A.R., Moriarty, E.C., 2004. The importance of proper model assumption in Bayesian phylogenetics. Syst. Biol. 53, 265–277.
Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452. doi:10.1093/bioinformatics/btp187.
Lin, C.-P., Danforth, B.N., 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. 30, 686–702.
Linares, M.C., Soto-Calderón, I.D., Lees, D.C., Anthony, N.M., 2009. High mitochondrial diversity in geographically widespread butterflies of Madagascar: a test of the DNA barcoding approach. Mol. Phylogenet. Evol. 50, 485–495.
Loader, S.P., Gower, D.J., Ngalason, W., Menegon, M., 2010. Three new species of *Callulina* (Amphibia: Anura: Brevicipitidae) highlight local endemism and conservation plight of Africás Eastern Arc forests. Zool. J. Linn. Soc. 160, 496–514.
Martens, J., Tietze, D.T., Eck, S., Veith, M., 2004. Radiation and species limits in the Asian Pallas's warbler complex (*Phylloscopus proregulus s.l.*). J. Ornithol. 145, 206–222.
Macey, J.R., Strasburg, J.L., Brisson, J.A., Vredenburg, V.T., Jennings, M., Larson, A., 2001. Molecular phylogenetics of Western North American frogs of the *Rana boylii* species group. Mol. Phylogenet. Evol. 19, 131–141.
Meier, R., Shiyang, K., Vaidya, G., Peter, K.L.Ng., 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst. Biol. 55 (5), 715–728.
Meier, R., Zhang, G., Ali, F., 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. Syst. Biol. 57 (5), 809–813.
Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3 (12), e422.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, New York.
Nishiumi, I., Kim, C.H., 2004. Little genetic differences between Korean and Japanese populations in songbirds. Natl. Sci. Mus. Monogr. 24, 279–286.
Nijman, V., Aliabadian, M., 2010. Performance of distance-based DNA barcoding in the molecular identification of Primates. C. R. Biologies. 333, 11–16.
Olsson, U., Alström, P., Sundberg, P., 2004. Non-monophyly of the avian genus *Seicercus* (Sylviidae) revealed by mitochondrial DNA. Zool. Scr. 33 (6), 501–510.
Olsson, U., Alström, P., Ericson, P.G.P., Sundberg, P., 2005. Non-monophyletic taxa and cryptic species – evidence from a molecular phylogeny of leaf-warblers (*Phylloscopus*, Aves). Mol. Phylogenet. Evol. 36, 261–276.
Olsson, U., Alström, P., Svensson, L., Aliabadian, M., Sundberg, P., 2010. The Lanius excubitor (Aves, Passeriformes) conundrum–taxonomic dilemma when molecular and non-molecular data tell different stories. Mol. Phylogenet. Evol. 55, 347–357.
Palma, R.E., Rodriguez-Serrano, E., Rivera-Milla, E., Hernandez, C.E., Salazar-Bravo, J., Carma, M.I., Belmar-Lucero, S., Gutierrez-Tapia, P., Zeballos, H., Yates, T.L., 2010. Phylogenetic relationships of the pygmy rice rats of the genus *Oligoryzomys* Bangs, 1900 (Rodentia: Sigmodontinae). Zool. J. Linn. Soc. Lond. 160, 551–566.
Parkin, D.T., Collinson, M., Helbig, A.J., Knox, A.G., Sangster, G., Svensson, L., 2004. Species limits in *Acrocephalus* and *Hippolais* warblers from the Western Palearctic. Brit. Birds 97, 276–299.
Penhallurick, J., Wink, M., 2004. Analysis of the taxonomy and nomenclature of the Procellariiformes based on complete nucleotide sequences of the mitochondrial cytochrome *b* gene. Emu 104, 125–147.
Petren, K., Grant, P.R., Keller, L.F., 2005. Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. Mol. Ecol. 14, 2943–2957.
Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55 (4), 595–609.
Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Posada, D., Crandall, K.A., 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50 (4), 580–601.
Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.
Price, T., 2010. The roles of time and ecology in the continental radiation of the Old World leaf warblers (*Phylloscopus* and *Seicercus*). Phil. Trans. R. Soc. B 365, 1749–1762.
Qu, Y., Ericson, P.G.P., Lei, F., Gebauer, A., Kaiser, M., Helbig, A.J., 2006. Molecular phylogenetic relationship of snow finch complex (genera *Montifringilla*, *Pyrgilauda*, and *Onychostruthus*) from the Tibetan plateau. Mol. Phylogenet. Evol. 40, 218–226.
Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcoding of life data system (www.barcodinglife.org). Mol. Ecol. Notes 7, 355–364, doi:10.1111/j.1471-8286.2006.01678.
Roe, A.D., Sperling, F.A.H., 2007. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. Mol. Phylogenet. Evol. 44, 325–345.
Round, P.D., Hansson, B., Pearson, D.J., Kennerley, P.R., Bensch, S., 2007. Lost and found: the enigmatic large-billed reed warbler *Acrocephalus orinus* rediscovered after 139 years. J. Avian Biol. 138, 133–138.
Rubinoff, D., Cameron, S., Will, K., 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J. Heredity 97 (6), 581–594.
Sangster, G., 2000. Genetic distance as a test of species boundaries in the Citril finch *Serinus citrinella*: a critique and taxonomic reinterpretation. Ibis 142, 487–490.
Semina, A.V., Polyakova, N.E., Brykov, Vl.A., 2007. Analysis of mitochondrial DNA: Taxonomic and phylogenetic relationships in two fish taxa (Pisces: Mugilidae and Cyprinidae). Biochemistry-Moscow 72 (12), 1349–1355.
Schenk, J.J., Hufford, L., 2010. Effects of substitution models on divergence time estimates: simulations and an empirical study of model uncertainty using cornales. Syst. Bot. 35 (3), 578–592.
Sullivan, J., Swofford, D.L., 1997. Are Guinea Pigs Rodents? The importance of adequate models in molecular phylogenetics. J. Mamm. Evol. 4, 77–86.
Svensson, L., Prys-Jones, R., Rasmussen, P.C., Olsson, U., 2008. Discovery of then new specimens of large-billed reed warbler *Acrocephalus orinus*, and new insights in its distributional range. J. Avian Biol. 39, 605–610.
Swofford, D.L., 2003. PAUP∗: Phylogenetic Analysis Using Parsimony (∗ and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Swofford, D.L., Olsen, G.P., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, second ed. Sinauer, Sunderland, MA., pp. 407–514.
Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Trewick, S.A., 2000. Mitochondrial DNA sequences support allozyme evidence for cryptic radiation of New Zealand Peripatoides (Onychophora). Mol. Ecol. 9, 269–281.
Trewick, S.A., 2008. DNA Barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). Cladistics 24, 240–254.
Vences, M., Thomas, M., Bonett, R.M., Vieitas, D.R., 2005. Deciphering amphibian diversity through DNA barcoding. Phil. Trans. R. Soc. B 360, 1859–1868.
Virgilio, M., Backeljau, T., Nevado, B., De Meyer, M., 2010. Comparative performances of DNA barcoding across insect orders. BMC Bioinformatics 11, 206.
Waters, J.M., Allibone, R.M., Wallis, G.P., 2006. Geological subsidence, river capture, and cladogenesis of galaxiid fish lineages in central New Zealand. Biol. J. Linn. Soc. 88, 367–376.
Waugh, J., 2007. DNA barcoding in animal species: progress, potential and pitfalls. BioEssays 29, 188–197.
Weckstein, J., Zink, R., Blackwell-Rago, R.C., Nelson, D.A., 2001. Anomalous variation in mitochondrial genomes of White-crowned (*Zonotrichia leucophrysa*) and Golden-crowned (*Z. atricapilla*) Sparrows: pseudogenes, hybridization, or incomplete lineage sorting? Auk 118 (1), 231–236.
Wesson, D.M., McLain, D.K., Oliver, J.H., Piesman, J., Collins, F.H., 1993. Investigation of the validity of species status of *Ixodes dammini* (Acari: Ixodidae) using rDNA. Proc. Natl. Acad. Sci. USA 90, 10221–10225.
Whitworth, T.L., Dawson, R.D., Magalon, H., Baudry, E., 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). Proc. R. Soc. B 274, 1731–1739.
Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist – a case study in blue butterflies. Front. Zool. 4, 8. doi:10.1186/1742-9994-4-8.
Wilcox, T.P., Hugg, L., Zeh, J.A., Zeh, D.W., 1997. Mitochondrial DNA sequencing reveals extreme genetic differentiation in a cryptic species complex on Neotropical Pseudoscorpions. Mol. Phylogenet. Evol. 7 (2), 208–216.
Woog, F., Wink, M., Rastegar-Pouyani, E., Gonzalez, J., Helm, B., 2008. Distinct taxonomic position of the Madagascar stonechat (*Saxicola torquatus sibilla*) revealed by nucleotide sequences of mitochondrial DNA. J. Ornithol. 149, 423–430.
Yang, Z., Goldman, N., Friday, A., 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Mol. Biol. Evol. 11, 316–324.
Zhang, S., Yang, L., Yang, X., Yang, J., 2007. Molecular phylogeny of the yuhinas (Sylviidae: *Yuhina*): a paraphyletic group of babblers including *Zosterops* and Philippine *Stachyris*. J. Ornithol. 148, 417–426.

New insights into family relationships within the avian superfamily Sylvioidea (Passeriformes) based on seven molecular markers

Silke Fregin[a§], Martin Haase[a], Urban Olsson[b], Per Alström[c]

[a]Vogelwarte Hiddensee, Zoological Institute and Museum, Ernst Moritz Arndt University of Greifswald, 17489 Greifswald, Germany. silke.fregin@uni-greifswald.de

[a]Vogelwarte Hiddensee, Zoological Institute and Museum, Ernst Moritz Arndt University of Greifswald, 17489 Greifswald, Germany. martin.haase@uni-greifswald.de

[b]Department of Zoology, University of Göteborg, Box 463, SE-405 30 Göteborg, Sweden. urban.olsson@zool.gu.se

[c]Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, SE-750 07 Uppsala, Sweden. per.alstrom@slu.se

[§]Corresponding Author

Abstract

Background

The circumscription of the avian superfamily Sylvioidea is a matter of long ongoing debate. While the overall inclusiveness has now been mostly agreed on and 20 families recognised, the phylogenetic relationships among the families are largely unknown. We here present a phylogenetic hypothesis for Sylvioidea based on one mitochondrial and six nuclear markers, in total ~6.3 kbp, for 79 ingroup species representing all currently recognised families and some species with uncertain affinities, making this the most comprehensive analysis of this taxon.

Results

The resolution, especially of the deeper nodes, is much improved compared to previous studies. However, many relationships among families remain uncertain and are in need of verification. Most

families themselves were very well supported based on the total data set and also by indels. Our data do not support the inclusion of Hylia in Cettiidae. The genera Scotocerca and Erythrocercus are the next relatives to Cettiidae, but separated by relatively long internodes. The families Paridae, Remizidae and Stenostiridae clustered among the outgroup taxa and not within Sylvioidea.

Conclusions

Although the phylogenetic position of *Hylia* is uncertain we tentatively support the recognition of the family Hyliidae Bates, 1930 for this genus and *Pholidornis*. We propose new family names for the genera *Scotocerca* and *Erythrocercus*, Scotocercidae and Erythrocercidae, respectively, rather than including these in Cettiidae. We recommend that Paridae, Remizidae and Stenostiridae are not included in Sylvioidea.

Background

The order Passeriformes, also called passerines or perching-birds, is the largest of the 40 orders within the class Aves, including ~60% of all ~10500 living bird species [1]. The passerines are divided into three major groups, with Acanthisittidae (New Zealand wrens) being sister to the two large parvorders oscines and suboscines [2-5]. Oscines, "true" songbirds, possess a complex syrinx, which enables them to perform complex songs, whereas suboscines do not have this characteristic [6,7]. Passerida, the largest groups within oscines, can only be delimited by an insertion of one amino acid in exon 3 of the c-myc gene [8], but no synapomorphic morphological character is known to define this taxon. Within Passerida, the superfamily Sylvioidea has proved difficult to delineate based on morphology, because of apparent multiple events of convergent evolution [e.g. 9–12]. Several of these studies found evidence that Sylvioidea *sensu* Sibley and Ahlquist [12] and Sibley and Monroe [13], which was based on DNA-DNA hybridization studies, was not monophyletic. Recently, Sylvioidea has gone through a profound rearrangement based on various sets of molecular sequence data [14–18]. These studies showed that several of the families and subfamilies established by Sibley and Ahlquist [12] were non-monophyletic.

The first comprehensive study of the whole superfamily was based on one nuclear and one mtDNA sequence [14]. They identified 10 well supported major clades, which they proposed be recognized at the family level. One of the consequences of their revision was a temporary loss of the family name Sylviidae, which was previously recognized as the largest family within Sylvioidea. As the type genus of Sylviidae Leach, 1820, *Sylvia*, was shown to be nested within the large Timaliidae Vigors and Horsfield, 1827 assemblage, it was suggested to suppress Sylviidae, following the principle of stability [9,14,19]. However, Sylviidae was re-established by Gelang et al. [17], to coexist as separate family

next to Timaliidae.

Table 1: Classification of Sylvioidea by Dickinson [20] and Gill and Donsker [1]. Superscript numbers in second column refer to numbers in fourth column, indicating new family affiliations. Genera in brackets give the name used by Dickinson [20], which have changed according to IOC World Bird Names [1]. Only genera used in this study are shown.

Dickinson (2003) [20]		Gill and Donsker (2011) [1]	
		1 Panuridae	*Panurus*
		2 Nicatoridae	*Nicator*
		3 Alaudidae	*Mirafra, Ammomanes, Alauda*
		4 Pycnonotidae	*Pycnonotus, Arizelocichla (Andropadus), Atimastillas (Chlorocichla), Phyllastrephus, Hypsipetes (Ixos),*
Hirundinidae	Hirundininae: *Hirundo*[5], *Delichon*[5]	**5 Hirundinidae**	*Hirundo, Delichon,*
		6 Pnoepygidae	*Pnoepyga*
		7 Macrosphenidae	*Melocichla, Sphenoeacus, Macrosphenus, Sylvietta, Cryptillas (Bradypterus)*
		8 Cettiidae	*Scotocerca, Erythrocercus, Tesia, Cettia, Abroscopus, Hylia*
Aegithalidae	*Aegithalos*[9], *Leptopoecile*[9], *Psaltriparus*[9]	**9 Aegithalidae**	*Aegithalos, Leptopoecile, Psaltriparus*
		10 Phylloscopidae	*Phylloscopus, Seicercus*
		11 Acrocephalidae	*Nesillas, Acrocephalus, Calamonastides (Chloropeta), Hippolais*
		12 Locustellidae	*Dromaeocercus, Megalurus, Bradypterus, Locustella*
		13 Donacobiidae	*Donacobius*
		14 Bernieridae	*Oxylabes, Bernieria, Hartertula, Thamnornis, Xanthomixis, Crossleyia,*
Alaudidae	*Mirafra*[3], *Ammomanes*[3], *Alauda*[3]		
Cisticolidae	*Cisticola*[15], *Scotocerca*[8], *Prinia*[15], *Spiloptila*[15], *Apalis*[15], *Hypergerus*[15], *Camaroptera*[15], *Calamonastes*[15]	**15 Cisticolidae**	*Cisticola, Prinia, Spiloptila, Apalis, Hypergerus, Camaroptera, Calamonastes, Orthotomus, Artisornis, Eremomela*
Genera incertae sedis	*Orthotomus*[15], *Artisornis*[15]		
Pycnonotidae	*Pycnonotus*[4], *Andropadus*[4], *Chlorocichla*[4], *Phyllastrephus*[4], *Ixos*[4]		
Genera incertae sedis	*Nicator*[2], *Erythrocercus*[18]		
Sylviidae	Megalurinae:*Megalurus*[12] Acrocephalinae:*Tesia*[8], *Cettia*[8], *Bradypterus*[12, 7], *Dromaeocercus*[12], *Nesillas*[11], *Melocichla*[7], *Sphenoeacus*[7], *Locustella*[12], *Acrocephalus*[11], *Hippolais*[11]		
Genera incertae sedis	*Macrosphenus*[17], *Hylia*[8], *Oxylabes*[14], *Bernieria*[14], *Hartertula*[14], *Thamnornis*[14], *Xanthomixis*[14], *Crossleyia*[14]		
	Phylloscopinae: *Phylloscopus*[10], *Seicercus*[10],		

Dickinson (2003) [20]		Gill and Donsker (2011) [1]	
	Abroscopus[8], *Eremomela*[15], *Sylvietta*[7] Sylviinae: *Sylvia*[19]		
Timaliidae	*Pellorneum*[17], *Illadopsis*[17], *Pseudoalcippe*[19], *Pnoepyga*[6], *Stachyris*[16], *Dumetia*[16], *Chrysomma*[19], *Chamaea*[19], *Turdoides*[18], *Garrluax*[18], *Alcippe*[19], *Phyllanthus*[18], *Yuhina*[20], *Erpornis**, *Panurus*[1], *Paradoxornis*[19]	**16 Timaliidae**	*Dumetia, Stachyris*
		17 Pellorneidae	*Illadopsis, Pellorneum*
		18 Leiothrichidae	*Phyllanthus, Turdoides, Trochalopteron (Garrulax)*
		19 Sylviidae	*Pseudoalcippe, Sylvia, Lioparus, Chrysomma, Chamaea, Sinosuthora (Paradoxornis)*
Genera incertae sedis	*Chaetops**		
Zosteropidae	*Zosterops*[20]	**20 Zosteropidae**	*Yuhina, Zosterops*

Sylvioidea now comprised 20 families containing in total more than 1200 species in 221 genera. Table 1 shows the latest printed classification by Dickinson [20] and the continuously updated IOC (International Ornithological Congress) World Bird Names List [1]. Their classification has taken all of the recent molecular advances into account. The most recent changes were that the monotypic genera *Panurus* and *Nicator* were raised to family level, Panuridae and Nicatoridae, respectively (cf. [11,14,16,18]; Macrosphenidae was proposed as family-name for the "Sphenoeacus group" (cf. [16,18]; the name Megaluridae was synonymized with Locustellidae, as the latter was found to have priority [21]; the family Pnoepygidae was proposed for the genus *Pnoepyga* [17]; the four subfamilies Timaliinae, Pellorneinae, Leiotrichinae and Zosteropinae recognized within Timaliidae [17] were all elevated to family rank; and *Scotocerca*, *Erythrocercus* and *Hylia* were tentatively included in Cettiidae (cf. [16,18,22-26].

Despite the numerous studies on large-scale relationships within Sylvioidea, the relationships among the families are still largely unresolved. We here present a multilocus analysis of one mitochondrial and six nuclear markers, ~6300 aligned basepairs for 79 species with the aim to clarify the phylogeny.

Results

Sequence statistics

The combined dataset comprised 6332 aligned basepairs of nucleotide sequence data, one mitochondrial and six nuclear markers. Percentage of parsimony informative sites were as follows: recombination activating gene 1 (RAG1) 34% (652/1934), fibrinogen beta chain (FGB) 36% (229/632),

glyceraldehyde-3-phosphate dehydrogenase (GAPDH) 38% (166/439), myoglobin (MB) 42% (319/765), ornithine decarboxylase 1 (ODC1) 45% (355/796), mtDNA cytochrome b (MT-CYB) 46% (531/1143), and lactate dehydrogenase B (LDHB) 47% (291/624).

GARLI-PART found the tree with the highest likelihood in 53 of all 100 runs, the next best tree was found in 27 of the runs. These trees differed only in the topology of the outgroup taxa. Thus, in 80 out of 100 inferences, GARLI-PART found the same topology within Sylvioidea, which was identical to the Bayesian inference (BI) 50% majority rule tree with respect to the relationships within Sylvioidea.

In the BI, 80/78% (combined/nuclear data) of the nodes were well supported (PP ≥0.95), 17/17% had PPs between 0.51 and 0.94, and only 2/5% of the nodes were unresolved. In the ML analyses, 61/50% of the nodes had support values ≥85%, 26/28% between 50% and 84%, and 13/22% <50%.

Phylogeny of Sylvioidea

The tree based on the complete dataset is shown in Fig. 1, and the tree based on the nuclear dataset is shown in Fig. 2, with the results from the single-locus analyses indicated in the latter figure. There is generally good agreement between these two trees. All families in Sylvioidea with more than one representative had PP 1.00 and ML bootstrap support 100%, except for Macrosphenidae and Cettiidae *sensu* Gill and Donsker [1] (Cettiidae *sensu* Alström et al. [14] had 1.00/100% support). The difficulties to delimit these families and the ongoing conflict in the deeper nodes is shown in Fig. 3.

Nicatoridae, Alaudidae and Panuridae were sister to all other sylvioid taxa (node 4), with PP 1.00 but lower ML bootstrap support. The sister relationship of Alaudidae and Panuridae was highly supported in the combined and nuclear analyses. Macrosphenidae was sister to the other sylvioid families (node 5), albeit less supported in the ML bootstrap analyses of the combined data set.

The remaining families were divided into two major clades, 6 and 11. Clade 6 consisted of Cisticolidae, Locustellidae, Bernieridae, Donacobiidae, Acrocephalidae, and Pnoepygidae. These relationships were mostly only supported by BI, although clade 8, containing Bernieridae, Donacobiidae and Locustellidae, was strongly supported by both BI and ML. The sister relationship of Donacobiidae and Bernieridae (node 9) was weakly supported in all analyses. The sister clade to Cisticolidae (7) had varying support in the combined and nuclear analyses.

The largest clade (11) was poorly supported, with a basal polytomy consisting of Hirundinidae, Pycnonotidae and a clade (12) containing the remaining families. Within clade 12, the strongly supported clade 13 comprised Zosteropidae, Timaliidae, Pellorneidae, and Leiothrichidae with Sylviidae as their common sister group. The relationships among the families in clade 14 were uncertain, and differed between the analysis of the complete dataset and the one based on only nuclear loci. The sister relationship between Leiothrichidae and Pellorneidae, only weakly supported

in the combined data set, was well (pp=0.94) supported in the nuclear data set, but not well by ML.

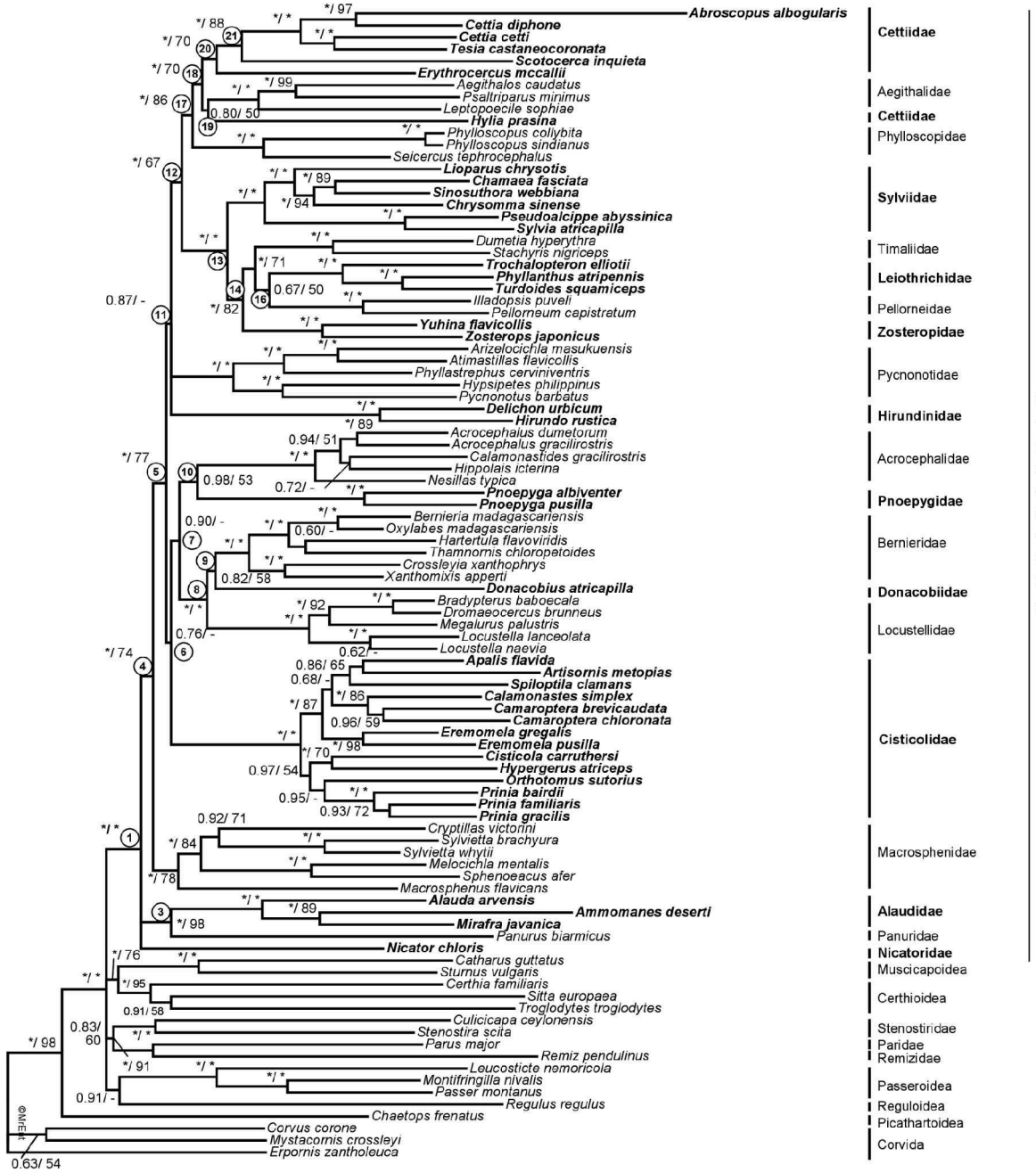

Figure 1: Sylvioidea phylogeny based on the complete data setPhylogenetic tree based on the complete concatenated dataset (MT-CYB, FGB, GAPDH, LDHB, MB, ODC1, RAG1), analysed by Bayesian inference. Support values are given in the order PP / ML bootstrap; an asterisk indicates PP=1.00 or ML=100%. For better clarity, families belonging to Sylvioidea are alternately written in bold. Node numbers are the same as in Fig. 2.

Clade 17 formed the sister clade to the sylviid/timaliid taxa (13), although the clade (12) containing these two clades received low ML bootstrap support. Within clade 17, Phylloscopidae was sister to a clade (19) containing Aegithalidae and a non-monophyletic Cettiidae. The sister relationship of Aegithalidae and the cettiid genus *Hylia* was poorly supported. The clade containing *Erythrocercus*,

Scotocerca and other Cettiidae (20) was well supported, especially by the nuclear data set, as was the *Scotocerca*/other Cettiidae clade (21).

There were only few strongly supported incongruences: 1) the sister relationship of *Ammomanes deserti* and *Mirafra javanica* (in Alaudidae) found by the complete and nuclear data sets, was strongly contradicted (PP 0.92–1.00) by the single-locus analyses of MB, GAPDH and MT-CYB, which instead

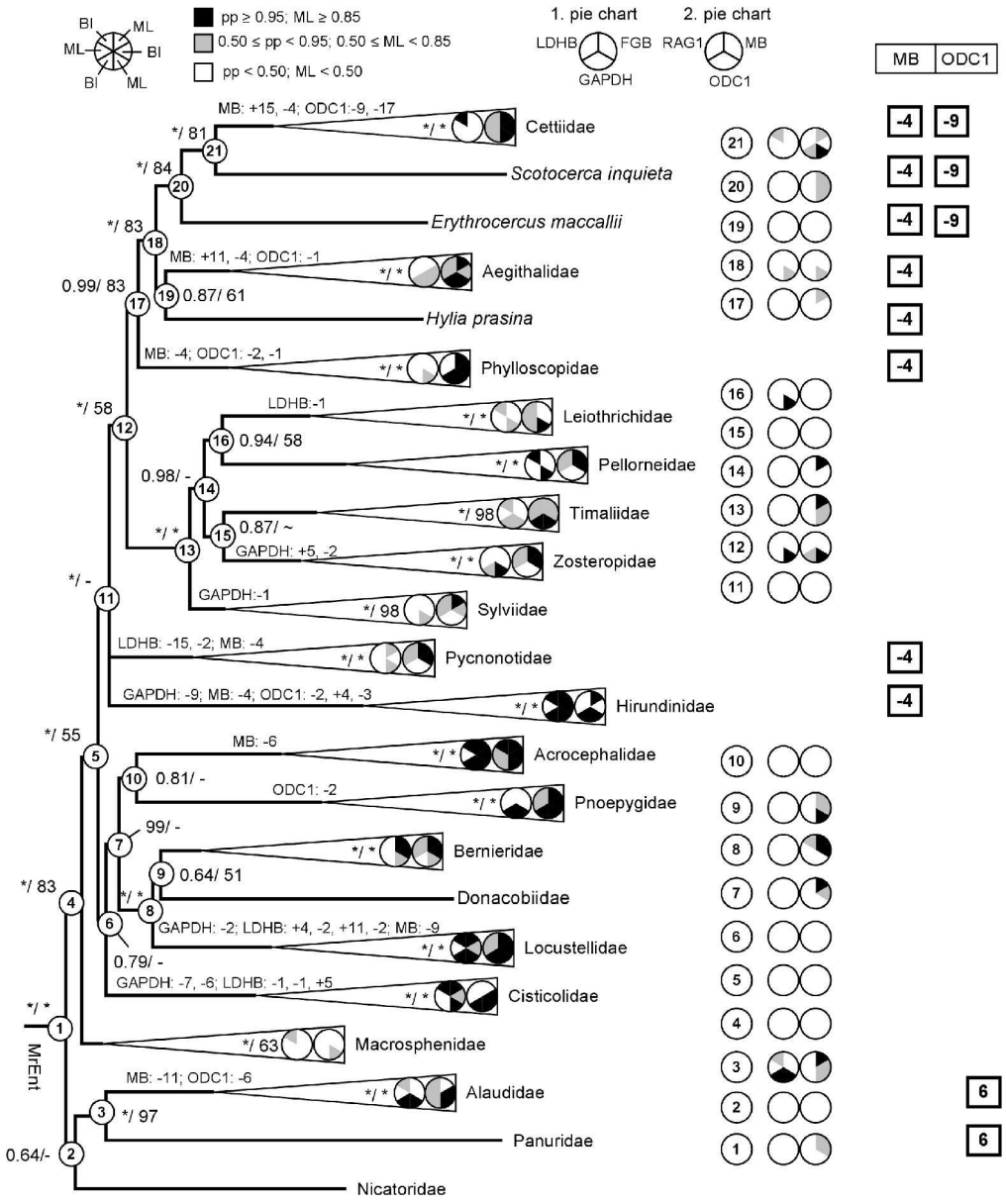

Figure 2: Sylvioidea phylogeny based on the nuclear data setPhylogenetic tree based on the nuclear dataset (FGB, GAPDH, LDHB, MB, ODC1, RAG1), analysed by Bayesian inference. Support values are given in the order PP / ML bootstrap (see also explanation in upper left corner of figure); an asterisk indicates PP=1.00 or ML=100%. Pie charts indicate support in the six nuclear single-locus analyses, first pie chart refers to FGB, GAPDH and LDHB; second pie chart refers to MB, ODC1 and RAG1. Pie charts in family clades indicate only support for the family itself, whereas pie charts on the right show support for nodes indicated on the tree. Squares on the right indicate indels shared among taxa in the alignments of the noncoding regions; only indels supporting more than one family or putative members of one family are shown. Numbers in squares indicate size of indels. Indels supporting individual families are given on the respective branches.

supported a sister relationship of *Alauda arvensis* and *Mirafra javanica*. 2) *Sinosuthora webbiana* was placed in Pellorneidae by FGB (PP=1.00). 3) *Donacobius* was sister to Locustellidae based on FGB, but sister to Bernieridae using ODC1. 4) *Trochalopteron elliotii* was placed in Pellorneidae and not in Leiothrichidae in the GAPDH tree (PP=1.00).

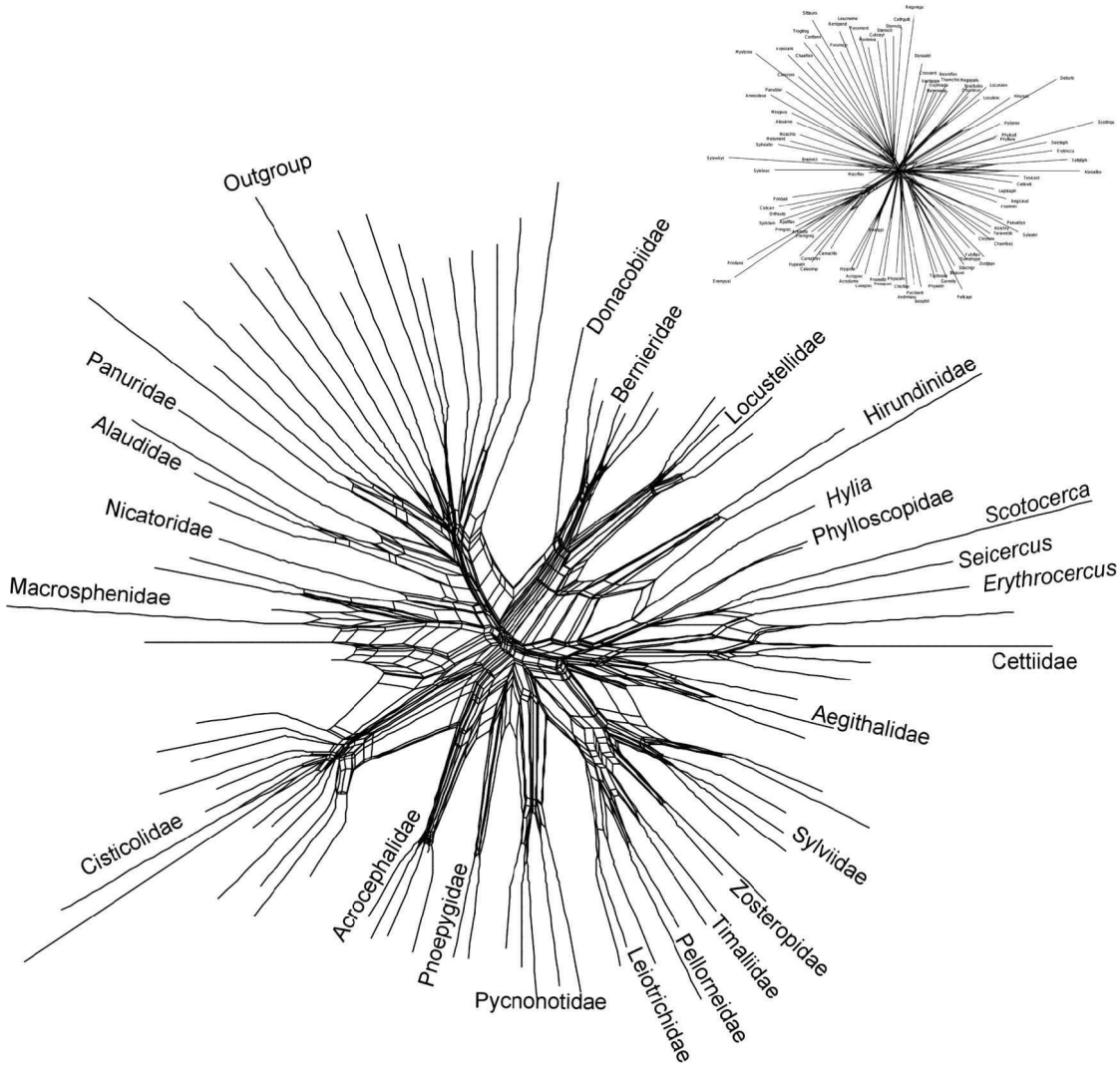

Figure 3: Network relationships within SylvioideaNeighborNet based on GTR+Γ distances. Centre of the network magnified, illustrating basal data conflict. Original network in upper right corner for comparison.

Indels

Most families had unique insertions and/or deletions (indels), which lent further support to these clades (Fig. 2).

However, few indels were shared by two or more families (Fig. 2). The grouping of Panuridae with Alaudidae was supported by an insertion of 6 bp in ODC1. *Erythrocercus* and *Scotocerca* shared a 9 bp deletion in ODC1 with the other Cettiidae, except *Hylia*. A 4 bp deletion in MB was shared by the taxa in clade 17 (Phylloscopidae, Aegithalidae and Cettiidae), but this was also found in Pycnonotidae and

Hirundinidae, which were inferred to be more distantly related. Two deletions of three basepairs in FGB and MB, respectively, delimited Sylvioidea from the outgroup, including Paridae, Remizidae and Stenostiridae. The inclusion of *Eremomela* in Cisticolidae was supported by several shared indels.

Discussion

Phylogeny of Sylvioidea

The present study is the most comprehensive analysis of the superfamily Sylvioidea, both with respect to the number of taxa and the number of loci. BI and ML searches found identical topologies, which reinforces the confidence in the results, even though the strength of the support differed between these methods. Only few deeper nodes (except those defining families) were supported by single-locus analyses. MB and ODC1 provided most resolution deep in the tree, and MB was the only single marker that supported Sylvioidea as a monophyletic group in the BI and ML bootstrap. The best ML trees for FGB and RAG1 also inferred Sylvioidea to be monophyletic, but this was not supported by their respective bootstrap analyses. Thus, the concatenation of all markers improved the resolution substantially.

The overall support of the multilocus tree, especially of the deeper nodes, had improved much compared to previous studies [14,16,18]. Especially studies using only mitochondrial data have failed to resolve most nodes below family level [27-29]. However, also an analysis by Johansson et al. [18] of a dataset comprising six loci (MB, ODC1, FGB, RAG1, RAG2 and ND2; in total ~7.3 kbp) for 14 sylvioid taxa was largely unresolved. The lack of resolution in the deeper nodes of the Sylvioidea phylogeny, probably due to the short internodes suggesting a rapid radiation of the families, was shown in the neighbor-net based on GTR-distances of Fig. 3. except Phylloscopidae and Cettiidae *sensu* Gill and Donsker [1], but not *sensu* Alström et al. [14], were well defined. Most conspicuous were the numerous possible pathways in the centre of the network, demonstrating the conflict between deeper splits in the Sylvioidea phylogeny.

The sister relationship of Alaudidae and Panuridae, which is extremely unexpected from a morphological and ecological perspective, was very well supported, also by several single-locus analyses. This relationship has been found also in previous studies based on fewer, but partly the same, loci [11,14,18,23]. The precise position of the enigmatic Nicatoridae still has to be regarded as uncertain.

The position of Macrosphenidae as sister to the remaining sylvioid taxa was well supported in the BI but less so in the ML bootstrap analyses. This was also found based on a different taxon sampling and partly different loci [16,18,21]. In contrast, in studies where only one mitochondrial and one nuclear

loci were used [14,23] Macrosphenidae was placed in a more derived position within Sylvioidea.

The two large clades 6 and 11 have been inferred in two previous studies based on different taxon sampling and some of the same loci as in the present analysis [21, 22], although they have not been recovered in other studies based on different taxon sampling and partly different loci [16,18]. As they were poorly supported here, they are to be considered as highly tentative.

Clade 7 in general was also found by various studies, but with differing constellations. While clade 8 was quite consistent between different studies [17,18,21 (Supplementary Fig. 2),22]; or excluding either Donacobiidae or Bernieridae: [14,16], the relationships between clade 8, Acrocephalidae and Pnoepygidae varied. The latter family was found as sister to clade 8 and Acrocephalidae [22,24] or in different positions [17], though never well supported. Lei et al. [27] found in their study based solely on mitochondrial sequences a close relationship between Locustellidae and Cisticolidae, but with Acrocephalidae falling in another clade, with high support in the Bayesian analysis, but with only low ML bootstrap support.

The largest clade (11) was divided into a polytomy formed by Pycnonotidae, Hirundinidae and clade 12. Pycnonotidae, Hirundinidae and clade 17 shared a 4 bp deletion in MB, which was not found in clade 13. Due to the somewhat uncertain relationships in the deeper nodes in this part of the tree, different scenarios are possible. One would be that this deletion was reversed by the members of clade 13, or that the different families lost these base pairs independently. Alternatively, the homoplastic appearance of this indel could also be a case of hemiplasy [31], were the gene tree is not congruent with the species tree due to incomplete lineage sorting. Hemiplasy is considered to be more likely when internodes are short, as it is the case in this clade. However, in a study about transposable elements over a wide range of birds cases of homoplasy were found, but lineage sorting was considered an unlikely explanation for these events [32].

Clade 12 was recovered also by Johansson et al. [18] (Fig. 2, clade I). Therein, clade 13 consisted of the much debated sylviid/timaliid families. All families had very high support in our study, as well as the whole clade (node 13) itself, which was only weakly supported in the ML analysis in Gelang et al. [17]. Based on a much denser taxon sampling but fewer loci than the present study the relationships among the families in this clade agreed with the latter study [17], although they were better supported. Sylviidae, when studied in larger sample sizes together with former Timaliidae [9,17,24] and based on more than one locus, were always found as a separate clade. Gelang et al. [17] recognised Leiothrichinae, Pellorneinae, Timaliinae and Zosteropinae as subfamilies within Timaliidae, whereas Gill and Donsker [1] elevated these to family rank. We support the latter treatment, as it is more on a par with the treatment of the other groups within Sylvioidea.

The close affinities of Phylloscopidae, Aegithalidae and Cettiidae (clade 17) were well supported by

our nuclear data set, although the relationships among these are not unanimously well supported by both BI and ML. This clade had previously been found [14,18,21,22], although with weaker support. The latter authors also noted morphological similarities between Cettiidae *sensu* Alström et al. [14], *Scotocerca*, *Erythrocercus* and Aegithalidae, especially between the first two (*Hylia* not examined).

The families Paridae, Remizidae and Stenostiridae are sometimes included in Sylvioidea [e.g. 13 (excluding Stenostiridae),33-35]. Based on the phylogeny presented here, additional evidence from indels, and previous studies, we recommend that these three families are not included in Sylvioidea, and accordingly that Sylvioidea is circumscribed as in Figs. 1 and 2.

Intrafamilial relationships

Macrosphenidae was the least supported family within Sylvioidea, and none of the single-locus analyses recovered this group with high support. This is probably the result of a long divergence between these species or species pairs, indicated by long branches. This clade contains species that are morphologically and ecologically highly divergent, and this in combination with some long internodes within this clade suggest that a number of extant and/or extinct taxa also belong here. In addition to the genera included here, also *Achaetops* has been shown to belong in this group [16].

Our results confirm the general structure within Cisticolidae recovered by Nguembock et al. [36]. We could also corroborate the sister relationship of *Calamonastes* and *Camaroptera*, which had previously been inferred based on single-locus analyses only [36,37]. Johansson et al. [18] suggested *Eremomela* to be nested within Cisticolidae, contra Dickinson [20] who placed it in Phylloscopinae. However, they found contradicting evidence in their study: ODC1 and MB supported a close relationship with *Apalis*, while FGB placed *Eremomela* as sister to *Prinia* (no other cisticolids were included). Our combined analyses placed *Eremomela* with high support in the clade including *Apalis*.

The present study included six out of the eight genera and six out of the eleven species in the Malagasy endemic Bernieridae, and is the most complete analysis of this family to date with respect to number of loci, although one mitochondrial study included three additional species (one additional genus: *Cryptosylvicola*) [29], and one study [21 (Supplementary Fig. 2)] based on MB, ODC1, LDH, GAPDH, and MT-CYB also included the monotypic genus *Cryptosylvicola*. All of the relationships inferred in the present study were strongly supported except for the sister relationship between *Hartertula* and *Thamnornis*.

Clade 18 consisted of Aegithalidae and Cettiidae, including the genera *Hylia*, *Erythrocercus* and *Scotocerca*, which have been assigned to Cettiidae [1,16,18,23,26]. Alström et al. [22] also noted that Cettiidae and *Scotocerca* shared certain morphological characters, such as 10 rectrices, whereas most passerines have 12. While *Erythrocercus* and *Scotocerca* were clearly related to Cettiidae *sensu* Alström

et al. [14] in the present study, a close affiliation of *Hylia* to Cettiidae is questionable. *Hylia* has proved to be difficult to place before [23,24,26], although Beresford et al. [16] found strong support for an unresolved *Hylia*/*Aegithalos*/*Cettia* clade based on the nuclear RAG1 and RAG2. However, strong support was found for a sister relationship between *Hylia* and *Pholidornis* based on mitochondrial ND2 and 12S [26]. The latter relationship has previously been suggested based on anatomical details [37], and *Hylia* and *Pholidornis* have been placed in the family Hyliidae [26,38]. This seems a reasonable treatment, although it would be desirable to include both *Hylia* and *Pholidornis* in a multilocus analysis, preferably including additional loci compared to the present study.

With respect to *Scotocerca*, we suggest that it is better placed in a monotypic family rather than in Cettiidae. It is morphologically and ecologically highly divergent from the Cettiidae *sensu* Alström et al. [14] (which admittedly is in itself a morphologically exceptionally variable group; cf. [39]). Moreover, it is separated from Cettiidae *sensu* Alström et al. [14] by a long internode, both in the present study and in the one by Alström et al. [22]. We therefore propose a new family name:

Scotocercidae Fregin, Haase, Olsson and Alström, new family-group name. Type genus *Scotocerca* Sundevall, 1872. Diagnosis: See del Hoyo et al. [40], pp. 465–466, and Plate 35, p.462.

We also suggest that the genus *Erythrocercus*, which includes three species distributed in sub-Saharan Africa, be treated as a monotypic family rather than in Cettiidae. The same reasons as for *Scotocerca* apply, although *Erythrocercus* is even more different morphologically [22]. We therefore propose a new family name:

Erythrocercidae Fregin, Haase, Olsson and Alström, new family-group name. Type genus *Erythrocercus* Hartlaub, 1857. Diagnosis: See del Hoyo et al. [40], pp. 327–328 and Plate 26, p. 324.

The familiy name Macrosphenidae for the "Sphenoeacus-group" is already widely used, but has not been formally described yet. Therefore, we introduce Macrosphenidae Fregin, Haase, Olsson and Alström. Type genus *Macrosphenus* Cassin, 1859. Diagnosis: See del Hoyo et al. [40], p. 641 and Plate 47, p. 640. This family includes the following genera: *Macrosphenus*, *Sphenoeacus*, *Melocichla*, *Achaetops*, *Sylvietta*, *Cryptillas*.

Materials and Methods

Taxonomy

Taxonomy follows the IOC World Bird Names List Version 2.10 July 2011 [1].

Taxon sampling and outgroup

We sampled 79 representatives of all 20 currently recognized families of the superfamily Sylvioidea

(Table 1, Appendix), represented by up to ten genera per family. We also included three species whose family affiliations are not fully understood yet: *Scotocerca inquieta*, *Erythrocercus mccallii,* and *Hylia prasina*.

The outgroup (Appendix) consisted of the three corvoid species *Erpornis zantholeuca*, *Mystacornis crossleyi* and *Corvus corone,* with which the tree was rooted; a close relative of Passerida (*Chaetops frenatus)*; two to three representatives from Passeroidea, Muscicapoidea, and Certhioidea; and representatives of Regulidae, Paridae, Remizidae and Stenostiridae.

DNA extraction, amplification, sequencing and assembly

DNA was extracted according to Miller et al. [41] with slight modifications or using the QIAamp® DNA MiniKit (50) following the manufacturer's protocol. The following loci were sequenced: the mitochondrial cytochrome *b* gene (MT-CYB; 1143 bp), the glyceraldehyde-3-phosphodehydrogenase intron 11 (GAPDH; 438 bp aligned), the complete nuclear lactate dehydrogenase intron 3 (LDHB; 624 bp aligned), the entire nuclear myoglobin intron 2 (MB; 765 bp aligned), the nuclear ornithine decarboxylase (ODC1) exon 6 (partial), intron 6, exon 7, intron 7 and exon 8 (partial) (in total 796 bp aligned), and a major part of the recombination-activating gene 1 (RAG1, 1934 bp). Not all loci were sequenced for all taxa (additional file 1). To reduce the risk of amplifying nuclear copies (numts) [42] in MT-CYB, this gene was amplified including flanking parts. PCRs were made up by single components or with Ready-To-GoTM PCR beads from GE Healthcare. PCR products were cleaned with ExoSap IT and products from cycle sequencing were cleaned with DyeEx 96Plate from Qiagen (only when the ABI sequencer was used). Sequencing was done on a LiCor DNA Sequencer Long READIR 4200 or on an ABI 3130xl Genetic Analyzer. Sequences were assembled manually in BioEdit [43] or with the Staden Package [44]. In addition, fibrinogen beta chain intron 5 sequences (FGB; 632bp aligned) were retrieved from GenBank. GenBank accession numbers for all included sequences are given in the additional file 1. All localities and museum sample numbers, which are new to this study are given in the Genbank record, respectively.

Phylogenetic analysis

The sequences were aligned using MAFFT [45] with complementary manual adjustments. Base compositions of the four different genetic markers were tested for nucleotide bias using č2 test of homogeneity across taxa implemented in PAUP * 4.0b10 [46]. All markers were tested for saturation effects with Dambe 5.2.34 [47,48]. No significant saturation was detected. Phylogenetic analyses were performed by Bayesian inference (BI) using MrBayes 3.1 [49,50] and maximum likelihood (ML)

inferences were conducted with GARLI-PART 0.97 [51]. Nine data sets were analysed: all seven loci separately, all concatenated (complete dataset), and all six nuclear loci concatenated (nuclear dataset). Indels were treated as missing data in BI and ML. In both multilocus analyses, the data were partitioned by locus, using rate multipliers to allow different rates for the different partitions.

MrModeltest [52] was used in conjunction with PAUP* [46] to estimate the best-fit nucleotide substitution models for implementation in MrBayes, based on the Akaike Information Criterion (AIC; [53]) and AICc for smaller samples [54,55]. The proposed models were: GTR+I+G for MB-CYB, GTR+G for FGB, HKY+G for GAPDH, GTR+G for LDHB, HKY+G for MB, JC for the exons of ODC1, GTR+G for the introns of ODC1 and GTR+I+G for RAG1. As GARLI-PART can implement more models than MrBayes, for the ML analyses jModelTest [56] was used to estimate nucleotide substitution models, with the same criteria as for MrModeltest. The best-fit models were: TVM+I+G for MT-CYB, TPM2uf+G for FGB, HKY+G for GAPDH, TPM3uf+G for LDHB, TPM3uf+G for MB, JC for the exons of ODC1, GTR+G for the introns of ODC1 and TIM3+I+G for RAG1. We conducted 100 ML search runs with GARLI-PART with random starting trees to obtain the tree with the maximum likelihood. Non-parametric bootstrapping was performed in GARLI-PART with 500 replicates for the combined, and 1000 replicates for single locus analyses. The resulting bootstrap trees were read into Treefinder version October 2008 [57,58] for obtaining the bootstrap values, as GARLI-PART does not calculateconsensus trees.

MrBayes was run with 4 to 8 chains for 10 to 21 million generation, in two parallel runs with default priors. In the single locus analyses of RAG1 temp=0.1 was used, as with default priors no convergence of both runs was obtained, even after several runs up to 30 million generations. Convergence of parameters in BI was monitored using the program Tracer v. 1.4 [59]. Burnin was defined as those number of generations that were obtained before the average standard deviation of split frequencies remained below 0.01. Thus, consensus trees were calculated from 40000 to 160000 trees, combined from both runs. We regard nodes with maximum likelihood bootstrap values >85% as well supported, following Erixon et al. [60], as it corresponds roughly to a 0.95 probability that the analyses recovered a correct clade, and posterior probabilities (PP) > 0.95.

A NeighborNet was calculated in SplitsTree4 version 4.11.2 [61] based on the best fit model (GTR+G) found for the complete data set by jModelTest [56], in order to illustrate data conflict in partitions at the base of the phylogeny.

Acknowledgements

SF is very grateful to the late AJ Helbig as initial supervisor and for initiating this study. SF is thankful to C Meibauer and A Kocum at the Vogelwarte Hiddensee, University of Greifswald for various

support and A Spillner for the introduction to using the Bioinformatic cluster, University of Greifswald. The molecular work was partly conducted at the Molecular Systematics Laboratory, Swedish Museum of Natural History, Stockholm, and SF is very thankful to M Irestedt and P Eldenäs for their support. We thank N David for comments on nomenclature and M Koopman for providing references. We are grateful to all sample collectors, too numerous to mention all individually.

References

1. Gill F, Donsker D: **IOC World Bird Names (version 2.10)**. [http://www.worldbirdnames.org/] (Accessed November 2011).
2. Barker FK, Barrowclough GF, Groth JG A: **phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data**. *Proc R Soc Lond B* 2002, **269**: 295–308.
3. Ericson PGP, Christidis L, Cooper A, Irestedt M, Jackson J, Johansson US, Norman JA: **A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens**. *Proc R Soc Lond B* 2002, **269**: 235–241.
4. Ericson PGP, Anderson CL, Britton T, Elzanowski A, Johansson US, Källersjö M, Ohlson JI, Parsons TJ, Zuccon D, Mayr G: **Diversification of Neoaves: integration of molecular sequence data and fossils**. *Biol Letters* 2006, **2**: 543–547.
5. Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T: **A phylogenomic study of birds reveals their evolutionary history**. *Science* 2008, **320**: 1763–1768.
6. Forbes WA: **Contributions to the anatomy of passerine birds Part 6 On *Xenicus* and *Achantisitta* as types of a new family (Xenicidae) of mesomyodian Passeres from New Zealand**. *Proc Zool Soc Lond* 1882, 569–571.
7. Ames PL: **The morphology of the syrinx in passerine birds.** *Bull Peabody Mus Nat Hist* 1971, **95**: 151–262.
8. Ericson PGP, Johansson US, Parson TJ: **Major divisions in oscines revealed by insertions in the nuclear gene c-myc: a novel gene in avian phylogenetics**. *Auk* 2000, **117**: 1069–1078.
9. Cibois A: **Mitochondrial DNA phylogeny of babblers (Timaliidae)**. *Auk* 2003, **120**: 35–54.
10. Cibois A, Pasquet E, Schulenberg TS: **Molecular systematics of the Malagasy babblers (Passeriformes: Timaliidae) and warblers (Passeriformes: Sylviidae) based on cytochrome b and 16S RNA sequences**. *Mol Phylogenet Evol* 1999, **13**: 581–595.
11. Ericson PGP, Johansson US: **Phylogeny of Passerida (Aves: Passeriformes) based on nuclear and mitochondrial sequence data**. *Mol Phylogenet Evol* 2003, **29**: 126–138.
12. Sibley CG, Ahlquist JE: *Phylogeny and Classification of Birds: a Study in Molecular Evolution*. New Haven CT: Yale University Press; 1990.
13. Sibley CG, Monroe BL Jr: *Distribution and Taxonomy of Birds of the World*. New Haven CT: Yale University Press; 1990.
14. Alström P, Ericson PGP, Olsson U, Sundberg P: **Phylogeny and classification of the avian superfamily Sylvioidea**. *Mol Phylogenet Evol* 2006, **38**: 381–397.
15. Barker FK, Cibois A, Schikler P, Feinstein J, Cracraft J: P**hylogeny and diversification of the largest avian radiation**. *PNAS* 2004, **101**: 11040–11045.

16. Beresford P, Barker FK, Ryan PG, Crowe TM: **African endemics span the tree of songbirds (Passeri): molecular systematics of several evolutionary 'enigmas'**. *Proc R Soc Lond B* 2005, **272**: 849–858.

17. Gelang M, Cibois A, Pasquet E, Olsson U, Alström P, Ericson PGP: **Phylogeny of babblers (Aves Passeriformes): major lineages family limits and classifications**. *Zool Scr* 2009, **38**: 225–236.

18. Johansson US, Fjeldså J, Bowie RCK: **Phylogenetic relationships within Passerida (Aves: Passeriformes): a review and a new molecular phylogeny based on three nuclear intron markers**. *Mol Phylogenet Evol* 2008, **48**: 858–876.

19. Cibois A: **Sylvia is a babbler: taxonomic implications for the families Sylviidae and Timaliidae**. *BOC* 2003, **123**: 257–261.

20. Dickinson EC: *The Howard and Moore Complete Checklist of the Birds of the World*. London: Christopher Helm; 2003.

21. Alström P, Fregin S, Norman JA, Ericson PGP, Christidis L, Olsson U: **Multilocus analysis of a taxonomically densely sampled dataset reveal extensive non-monophyly in the avian family Locustellidae**. *Mol Phylogenet Evol* 2011, **58**: 513–526.

22. Alström P, Fjeldså J, Fregin S, Olsson U: **Gross morphology betrays phylogeny: the Scrub Warbler *Scotocerca inquieta* is not a cisticolid**. *Ibis* 2011, **153**: 87–97.

23. Fuchs J, Fjeldså J, Bowie RCK, Voelker G, Pasquet E: **The African warbler genus *Hyliota* as a lost lineage in the oscine songbird tree: molecular support for an African origin of the Passerida**. *Mol Phylogenet Evol* 2006, **39**: 186–197.

24. Irestedt M, Gelang M, Sangster G, Olsson U, Ericson PGP, Alström P: **Neumann's Warbler *Hemitesia neumanni* (Sylvioidea): the sole African member of a Paleotropic Miocene avifauna**. *Ibis* 2011, **153**: 78–86.

25. Pasquet E, Cibois A, Ballon F, Érard C: **What are African monarchs (Aves Passeriformes)? A phylogenetic analysis of mitochondrial genes**. *CR Biol* 2002, **325**: 1–12.

26. Sefc KM, Payne RB, Sorenson MD: **Phylogenetic relationships of African sunbird-like warblers: Moho (*Hypergerus atriceps*) Green Hylia (*Hylia prasina*) and Tit-hylia (*Pholidornis rushiae*)**. *Ostrich* 2003, **74**: 8–17.

27. Lei X, Yin Z, Lian Z, Chen C, Dai C, Kristin A, Lei F: **Phylogenetic relationships of some Sylviidae species based on complete mtDNA cyt b and partial COI sequence data**. *Chinese Birds* 2010, **1**: 175–187.

28. Barhoum DN, Burns KJ: **Phylogenetic relationships of the Wrentit based on mitochondrial cytochrome b sequences**. *Condor* 2002, **104**: 740–749.

29. Cibois A, Slikas B, Schulenberg T S, Pasquet E: **An endemic radiation of Malagasy songbirds is revealed by mitochondrial DNA sequence data.** *Evolution* 2001, **55**: 1198–1206.

30. Wolters HE: *Die Vögel Europas im System der Vögel: Eine Übersicht*. Bonn. Baden-Baden: Biotropic Verlag; 1983.

31. Avise JC, Robinson TJ: **Hemiplasy: a new term in the lexicon of phylogenetics**. *Syst Biol* 2008, **57**: 503–507.

32. Han K-L, Braun EL, Kimball RT, Reddy S, Bowie RCK, Braun MJ, Chojnowski JL, Hackett SJ, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T: **Are Transposable Element Insertions Homoplasy Free?: An Examination Using the Avian Tree of Life**. *Syst Biol* 2011, **60**: 375–386.

33. Cracraft J, Barker FK, Braun M, Harshman J, Dyke GJ, Feinstein J, Stanley S, Cibois A, Schikler P, Beresford P, García-Moreno J, Sorenson MP, Yuri T, Mindell DP: **Phylogenetic**

relationships among modern birds (Neornithes) Toward an avian tree of life. In *Assembling the Tree of Life.* Edited *by* Cracraft J Donoghue MJ. Oxford: Oxford University Press; 2004: 468–489.

34. Harshman John: The Tree of Life Web Project http://tolweborg/Sylvioidea/67276/20060802 (Sylvioidea Version 02 August 2006; under construction)
35. Sangster G, Collinson JM, Knox AG, Parkin DT, Svensson L: **Taxonomic recommendations for British birds: Sixth report**. *Ibis* 2010, **152**: 180–186.
36. Nguembock B, Fjeldså J, Tillier A, Pasquet E: **A phylogeny for the Cisticolidae (Aves: Passeriformes) based on nuclear and mitochondrial DNA sequence data and a re-interpretation of an unique nest-building specialization**. *Mol Phylogenet Evol* 2007, **42**: 272–286.
37. Nguembock B, Fjeldså J, Couloux A, Cruaud C, Pasquet E: **Polyphyly of the genus *Apalis* and a new generic name for the species *pulchra* and *ruwenzorii***. *Ibis* 2008, **150**: 756–765.
38. Bates GL: *Handbook of the birds of West Afric*a. London: Bale Sons and Danielson; 1930.
39. Alström P, Höhna S, Gelang M, Ericson PGP, Olsson U: **Non-monophyly and intricate morphological evolution within the avian family Cettiidae revealed by multilocus analysis of a taxonomically densely sampled dataset**. *BMC Evol Biol* 2011, **11**: 352 doi:10.1186/1471-2148-11-352.
40. del Hoyo J Elliott A Sargatal J: *Handbook of the Birds of the World*: Vol 11. Barcelona: Lynx Edicions; 2006: 378–490.
41. Miller SA, Dykes DD, Polesky HF: **A simple salting out procedure for extracting DNA from human nucleated cells**. *Nucleic Acids Res* 1988, **16**: 1215.
42. Sorenson MD, Quinn TW: **Numts: a challenge for avian systematics and population biology**. *Auk* 1998, **115**: 214–221.
43. Hall TA: **BioEdit: a user–friendly biological sequence alignment editor and analysis program for Windows 95/98/NT**. *Nucl Acids Symp Ser* 1999, **41**: 95–98.
44. Bonfield JK, Smith KF, Staden R: **A new DNA sequence assembly program**. *Nucl Acids Res* 1995, **24**: 4992–4999.
45. Katho K, Misawa K, Kuma K-i, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform**. *Nucl Acids Res* 2002, **30**: 3059–3066.
46. Swofford DL: *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4:* Sinauer Associates Sunderland Massachusetts; 2003.
47. Xia X, Xie Z, Salemi M,, Chen L, Wang Y: **An index of substitution saturation and its application**. *Mol Phylogenet Evol* 2003, **26**: 1–7.
48. Xia X, Lemey P: **Assessing substitution saturation with DAMBE.** In *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny.* 2nd edition. Edited by Lemey P, Salemi M, Vandamme A-M. Cambridge University Press; 2009:615–630.
49. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny.** *Bioinformatics* 2001, **17**: 754–755.
50. Ronquist F, Huelsenbeck JP: **MRBAYES 3: Bayesian phylogenetic inference under mixed models**. *Bioinformatics* 2003, **19**: 1572–1574.
51. Zwickl DJ: **Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion**. *PhD thesis.* The University of Texas at Austin; 2006.
52. Nylander JAA: MrModeltest v2 Program distributed by the author Evolutionary Biology Centre Uppsala University. [http://wwwabcse/~nylander/]

53. Akaike H: I**nformation theory as an extension of the maximum likelihood principle**. In *Second International Symposium on Information Theory:* 1973; Budapest. Edited by Petrov BN and Csaki F: Akademiai Kiado 1973:267–281.
54. Hurvich CM, Tsai C-L: **Regression and time series model selection in small samples.** *Biometrika* 1989, **76**: 297–307.
55. Sugiura N: **Further analysis of the data by Akaike's information criterion and the finite corrections**. *Commun Stat Theory Methods* 1978, **A7**:13–26.
56. Posada D: **jModelTest: phylogenetic model averaging**. *Mol Biol Evol* 2008, **25**: 1253–1256.
57. Jobb G, von Haeseler A, Strimmer K: **TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics**. *BMC Evol Biol* 2004, **4**: 1471–2148.
58. Jobb G: TREEFINDER version of October 2008 Munich Germany Distributed by the author at [wwwtreefinderde]
59. Rambaut A, Drummond AJ: Tracer v1.4 Available from [http://beastbioedacuk/Tracer]
60. Erixon P, Svennblad B, Britton T, Oxelman B: **Reliability of Bayesian Posterior probabilities and bootstrap frequencies in phylogenetics**. *Syst Biol* 2003, **52**: 665–673.
61. Huson DH, Bryant D: **Application of Phylogenetic Networks in Evolutionary Studies**. *Mol Biol Evol* 2006, **23**: 254-267.

Additional file 1

List of samples (in alphabetical order), with GenBank accession numbers

	MT-CYB	FGB	GAPDH	LDHB	MB	ODC1	RAG1
Ingroup							
Abroscopus albogularis	HQ706175	–	HQ706264	HQ706186	HQ706226	HQ706303	
Acrocephalus dumetorum	AJ004773	EF626749	FJ357911	FJ883062	FJ883105	FJ883134	FJ358146
Acrocephalus gracilirostris	AJ004782	–	–	FJ883063	FJ883095	FJ883135	
Aegithalos caudatus		EU680621	FJ357912		AY228281	EU680703	
Alauda arvensis	–	EF626747	FJ357913	HQ333047	AY228284	EF625336	AY056978
Ammomanes deserti		–					
Apalis flavida	HQ333036	–	HQ333097	HQ333049	HQ333069	HQ333083	
Arizelocichla masukuensis	HQ333035	EF626698	–	HQ333048		EF625287	
Artisornis metopias		EU686247					
Atimastillas flavicollis		EF626721				EF625310	
Bernieria madagascariensis	HQ333038	–	HQ333100	HQ333052	HQ333071	HQ333086	
Bradypterus baboecala	FJ883053	–	HQ333098	FJ883090	DQ008525	FJ883162	
Calamonastes simplex		–				–	
Calamonastides gracilirostris	FJ883043	–	–	FJ883077	FJ883113	FJ883149	
Camaroptera brevicaudata		–	–				
Camaroptera chloronota		–					
Cettia cetti			HQ706263	HQ333053	HQ706225	HQ121555	
Cettia diphone	DQ008510	EU680634	HQ121536	–	EU680584	EU680714	–
Chamaea fasciata	AJ534526	–	FJ357856		FJ357927	FJ358025	FJ358091
Chrysomma sinense		–	FJ357857		FJ357928	FJ358026	FJ358092
Cisticola carruthersi		–				–	
Crossleyia xanthophrys	HQ706177	–	HQ706269	HQ706191	HQ706231	HQ706309	
Cryptillas victorini	AY958231	EU680630	–	–	EU680602	EU680710	AY799815
Delichon urbicum	DQ008517	EU680641	HQ333103	HQ333055	DQ008568	EU680721	–
Donacobius atricapillus	DQ008481	EU680643	FJ357915	HQ333054	DQ008533	EU680723	AY319979
Dromaeocercus brunneus	HQ706160	EU680644	HQ706289	HQ706211		EU680724	
Dumetia hyperythra		–	FJ357859		FJ357930	FJ358028	FJ358094
Eremomela gregalis		EU680646	–			EU680726	
Eremomela pusilla		–					–
Erythrocercus mccallii	AF096465	EU680647	HQ121544	–	EU680585	EU680727	–
Hartertula flavoviridis	HQ706131	EU680650	HQ706267	HQ706189	HQ706229	HQ706307	
Hippolais icterina	DQ008479	EU680651	HQ333104	FJ883078	FJ883120	FJ883153	
Hirundo rustica		EF626748		HQ333056	AY064258	EF441240	–
Hylia prasina	HQ333041	EU680652	HQ121545	HQ333057	EU680583	EU680732	AY319984
Hypergerus atriceps		–	–				
Hypsipetes philippinus		EF626742	–			EF625331	
Illadopsis puveli		EU686236	–			–	
Leptopoecile sophiae		EU680658	HQ706262		DQ008569	EU680738	
Lioparus chrysotis		–	FJ357874			FJ358043	
Locustella lanceolata	HQ706139	–	HQ706275		HQ706235	HQ706313	
Locustella naevia	HQ706147	–	HQ706282		HQ706242	HQ706320	EF568259
Macrosphenus flavicans	–	EF626751	–	–	EF625286	EF625340	AY319987
Megalurus palustris	FJ883052	EU680661		FJ883089	DQ008529	FJ883161	
Melocichla mentalis	DQ008500	–	HQ333107	HQ333059	DQ008551	HQ333090	
Mirafra javanica	DQ008520	–	HQ333106	HQ333058	DQ008571	HQ333089	
Nesillas typica	–	EU680665	–	–	EU680592	EU680744	–
Nicator chloris		EU680666			EU680603		AY319991
Orthotomus sutorius	DQ008491	GQ242050	HQ333109	HQ333061	DQ008542	HQ333092	AY319992

	MT-CYB	FGB	GAPDH	LDHB	MB	ODC1	RAG1
Oxylabes madagascariensis	HQ706179	–	HQ706266	HQ706188	HQ706228	HQ706306	
Panurus biarmicus		EU680668	FJ357919		FJ357983	FJ358083	FJ358150
Pellorneum capistratum		–				–	
Phyllanthus atripennis		–	FJ357888			FJ358057	
Phyllastrephus cerviniventris		EF626726	–			EF625315	
Phylloscopus collybita	Z73487	GQ242051	FJ357920	–	DQ125966	FJ358084	AY319997
Phylloscopus sindianus	Z73478	–	HQ706302	HQ706224	HQ706261	HQ706340	
Pnoepyga albiventer	HQ121521	–	FJ357889	–	FJ357959	FJ358058	FJ358124
Pnoepyga pusilla		–	FJ357890	–	FJ357960	FJ358059	FJ358125
Prinia bairdii	AY352536	EU680675	FJ357921	–			FJ358151
Prinia familiaris	DQ008490	–	HQ121547	HQ333063	DQ008541	HQ121557	–
Prinia gracilis		–	–				
Psaltriparus minimus	GU244418	EU680678	–	–	EU680582	EU680757	AY319999
Pseudoalcippe abyssinica	AJ534548	EU680679	–			EU680758	
Pycnonotus barbatus	HQ333043	EF626746	HQ333110	HQ333062	HQ333075	HQ333093	FJ358152
Scotocerca inquieta	HQ333044	–	HQ333111	HQ333064	HQ333076	HQ333094	–
Seicercus tephrocephalus	HQ706182	–	HQ706301	HQ706223	HQ706260	HQ706339	–
Sinosuthora webbiana		EU680669				EU680748	
Sphenoeacus afer		EU680687	–	HQ333066		EU680766	AY799822
Spiloptila clamans	DQ008495	–	–		DQ008546		
Stachyris nigriceps	HQ333045	EU680688	FJ357900	HQ333065	FJ357969	HQ333095	FJ358135
Sylvia atricapilla	Z73494	EU680691	EF441232	HQ333067	AY887727	EF441254	EF568261
Sylvietta brachyura		–	–		DQ125960	–	
Sylvietta whytii	DQ008501	EU680693	HQ121548	–	DQ008552	EU680772	–
Tesia castaneocoronata	JN808933	–		–	JN809081	JN809044	
Thamnornis chloropetoides	HQ333046	EU680694	FJ357923	HQ333068	HQ333077	HQ333096	AY320004
Trochalopteron elliotii		–					
Turdoides squamiceps		–					
Xanthomixis apperti	HQ706181	–	HQ706265	HQ706187	HQ706227	HQ706305	
Yuhina flavicollis		–	FJ357908		FJ357977	FJ358077	FJ358143
Zosterops japonicus	AB159168	–	FJ357910	–	FJ357979	FJ358079	FJ358145
Outgroup							
Catharus guttatus	EU619718	EU680632	–	–	DQ466820	EU680712	AY307184
Certhia familiaris		EU680633	–		DQ011861	EU680713	AY056983
Chaetops frenatus	AY228052	EU680635	EF441212		AY228289	EU680715	AY443266
Corvus corone	U86032	AY529982	FJ357914	–	FJ357980	EU272116	AY056989
Culicicapa ceylonensis	AF096453	EU680640	GQ369627		EU680605	EU680720	AY443279
Erpornis zantholeuca		–	–		–	–	AY443339
Leucosticte nemoricola		–	–				
Montifringilla nivalis		–	–		DQ244066		
Mystacornis crossleyia		FJ178360					
Parus major	EU167009	EU680670	EU272098		AY228310	EU680749	AY443314
Passer montanus	AY030118	EF626752	AY336586	–	AY228311	EU325847	AF143738
Regulus regulus	AJ004762	EU680682	–		DQ008572	EU680761	
Remiz pendulinus	AY228081	EU680683	–		AY228319	EU680762	AY443328
Sitta europea	AF378102	EU680686	–		AY064257	EU680765	AY064272
Stenostira scita		EU680689	GQ369629		EU680607	EU680768	AY799823
Sturnus vulgaris	HM633385	EU680690	–		AY228322	EU680769	DQ466812
Troglodytes troglodytes		EU680696	–		AY228325	EU680775	

Danksagung

Ich bin Andreas J. Helbig sehr dankbar, dass er mir die Chance gegeben hat, diese Doktorarbeit unter seiner Leitung zu beginnen. Leider verstarb er viel zu früh und hat die Ergebnisse und die Fertigstellung der Arbeit nicht miterleben können. Ich hoffe aber, dass sie in seinem Sinne sind.

Etwa ein Jahr später lernte ich, dank Dorit Liebers-Helbig, Per Alström aus Uppsala, Schweden, auf dem „International Ornithological Congress“ in Hamburg kennen. Nach einem kurzen Gespräch wurde ich spontan von Per „adoptiert“. Vielen herzlichen Dank für Deine Unterstützung, Zusammenarbeit und Führung durch den „Warbler-Dschungel“!!

Nachdem Angela Schmitz Ornés und Martin Haase die neuen Leiter der Vogelwarte wurden, fand nach ein paar Monaten der Umzug der Vogelwarte Hiddensee nach Greifswald statt. Nun hatte ich auch „Vor-Ort-Hilfe“. Ich bedanke mich sehr für Eure Unterstützung und Geduld! Ein großes Dankeschön geht besonders an Martin, u.a. für die Diskussionen und Hilfe bezüglich der Analysen.

Michael Weiß aus Tübingen hat Annett und mich in die Welt der Phylogenieprogramme eingeführt. Angefangen von der Bedienung von Paup auf einem Macintosh-Computer bis hin zu „batch-files“ für MrBayes. Außerdem hat er es uns ermöglicht, an seinem Phylogeniekurs über zwei Wochen an der Universität von Tübingen teilzunehmen. Während dieser Zeit wurden wir herzlich von ihm und seiner Familie in seinem Haus aufgenommen. Vielen herzlichen Dank!

Die Arbeit für einen großen Teil meiner Sequenzdaten konnte ich im Labor für molekulare Sytematik am Naturhistorischen Museum in Stockholm, Schweden, durchführen. Hier danke ich besonders Pia Eldenäs und Martin Irestedt für ihre stets freundliche Unterstützung! Mein Dank gilt auch dem Synthesys-Programm, welches den Forschungsaufenthalt finanziert hat (SE-TAF-2992).

Einen wichtigen Anteil an dieser Arbeit hatten auch alle, die Proben gesammelt, mir DNS oder Sequenzen überlassen haben: P. Alström, H. Ansorge, F. Bairlein, H. Bartsch, A. Berruti, L. DeBrooke, D. Brown, C. Cicero, J. Dierschke, J. Fjeldså, W. Forstmeyer, D. Franz, K. Gauger, M. Gelang, B. Giessing, H. Grimm, G. Halmos, A.J. Helbig, U. Johansson, A. Kocum, J.B. Kristensen, B. Leisler, J. Martens, E. Mey, T. Michaelson, J. Morgan, G. Nikolaus, U. Olsson, E. Pasquet, D.J. Pearson, R. Pfeifer, T. Price, S. Rumsey, P. Round, V. Salewski, H. Shirihai, C. Spottiswoode, M. Stubbe, L. Svensson, E. Thaler, F. Wong. Vielen Dank!

Andreas Spillner danke ich für die Einweisung in die Bedienung des Computerclusters der Bioinformatik in Greifswald.

Vielen Dank an Lynx Edicions (Handbook of the Birds of the World), Paul Vinke und Micha Luhn für die Erlaubnis ihre Bilder zu verwenden.

Eine sehr wichtige Rolle haben meine „Mitbewohner“ Annett Kocum, Viviane Sternkopf, Christel Meibauer und Hausi Olaf Milski gespielt. Allen voran gilt mein Dank besonders Annett: sie hat mich in die Labor-Welt eingeführt und ist zwischen Fototouren über die Insel und Nachtwanderungen zum Leuchtturm zu einer sehr guten Freundin geworden! Christel möchte ich besonders für ihre Hilfe beim Sortieren und Beschriften von kryptischen Proben aus den Tiefen der Gefriertruhe danken! Und ohne Hausi, das Versorgungs-Kernstück der Vogelwarte auf Hiddensee, hätte sowieso nichts funktioniert. Euch allen danke ich für die tolle WG-Zeit!

Ein herzliches Dankeschön geht auch an Susan Zielske und Henriette Höltje für erholsame Kreativpausen und fürs Korrekturlesen!

Last but not least geht mein Dank an Micha: für Deine komplett unsylvioidaeische und dafür umso wichtigere Unterstützung und Geduld!