Institut für Biochemie und Biologie

Evolutionsbiologie/Spezielle Zoologie

Inference of phylogenetic relationships in passerine birds (Aves: Passeriformes) using new molecular markers

Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin "Evolutionsbiologie"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von
Simone Treplin

Potsdam, August 2006

Acknowledgements

First of all, I would like to thank Prof. Dr. Ralph Tiedemann for the exciting topic of my thesis. I'm grateful for his ongoing interest, discussions, support, and confidence in the project and me.

I thank the University of Potsdam for the opportunity to perform my PhD and the financial and logistical funds.

This thesis would not have been possible without many institutions and people, who provided samples: University of Kiel, Haustierkunde (Heiner Luttmann and Joachim Oesert), Zoologischer Garten Berlin (Rudolf Reinhard), Tierpark Berlin (Martin Kaiser), Transvaal Museum, South Africa (Tamar Cassidy), Vogelpark Walsrode (Bernd Marcordes), Eberhard Curio, Roger Fotso, Tomek Janiszewski, Hazell Shokellu Thompson, and Dieter Wallschläger. Additionally, I thank everybody who thought of me in the moment of finding a bird, collected and delivered it immediately.

I express my gratitude to Christoph Bleidorn for his great help with the phylogenetic analyses, the fight with the cluster, the discussions, and proof-reading. Special thanks go to Susanne Hauswaldt for patiently reading my thesis and improving my English.

I thank my colleagues of the whole group of evolutionary biology/systematic zoology for the friendly and positive working atmosphere, the funny lunch brakes, and the favours in the lab. I'm grateful to Romy for being my first, 'easy-care' diploma-student and producing many data.

I'm very grateful to Katja for her patience and perpetual technical help.

A very special thank goes to Philine for her friendship, encouragement, understanding, and good ideas. It would have been harder without her.

Thanks to Steffi for cat-sitting on Potsdam-weekends and together with 'the Henne' for realistic birds. I hope you understand why I preferred Mama's nice painting of the fieldfare for the cover.

I could not have reached anything without Malte and my parents, who have supported me throughout the years and are there for me whenever I need them. Thank you for love, confidence, encouragement, and being with me.

Table of contents

1 Introduction

1.1 Phylogenetic relationships within Passeriformes and the need for new markers

Among all classes of living organisms, Aves is supposed to be the best known, and some argue that presumably 'all' species have been discovered and named (Groth and Barrowclough, 1999). Nevertheless, their origin, phylogeny, and biogeography has been a continuous matter of debate, which has been intensified through the use of molecular data (e.g. Cracraft, 2001; Groth and Barrowclough, 1999; Sibley and Ahlquist, 1990). The difficulty in resolving these issues stems from their rapid adaptive radiation and the adaptation to flight. The anatomical characteristics correlated with the development of flight gained by the first birds are more or less conserved in recent species and thus, birds own only few taxon specific morphological synapomorphies (Feduccia, 1996).

The highest diversity among living birds is found in the order Passeriformes. This by far largest avian taxon comprises roughly 59 % of all living birds (more than 5700 species, Sibley and Ahlquist, 1990). The Passeriformes form a morphologically very homogenous group and their monophyly is well established, both on morphological (Raikow, 1982) and molecular grounds (Sibley and Ahlquist, 1990). However, phylogenetic relationships within the group have been extremely puzzling, as most of the evolutionary lineages originated through rapid radiation during the early Tertiary (Feduccia, 1995). Fast diverging clades had little opportunity to acquire synapomorphies, which resulted in ill-defined groups for reconstructions of a phylogeny (Lanyon, 1988).

The first extensive molecular study on avian systematics was based on DNA-DNA hybridization analyses (Sibley and Ahlquist, 1990) and corroborated the basal split of Passeriformes into the two morphologically monophyletic clades of suboscines (Tyranni) and oscines (Passeri) (e.g. Ames, 1971; Feduccia, 1975). This study, however, has been criticised by several authors concerning its reproducibility (Mindell, 1992), sparse sampling and its lack of internal consistency (Cracraft, 1992; Lanyon, 1992). Nevertheless, Sibley and Ahlquist's (1990) phylogeny of the Passeriformes (Fig. 1) with 46 families and 46 subfamilies (classified by Sibley and Monroe (1990)) has become the basis for subsequent DNA sequence analyses. While sequence-based studies generally agree with the partition of Passeriformes into the monophyletic clades of suboscines and oscines, a third group composed of the New Zealand

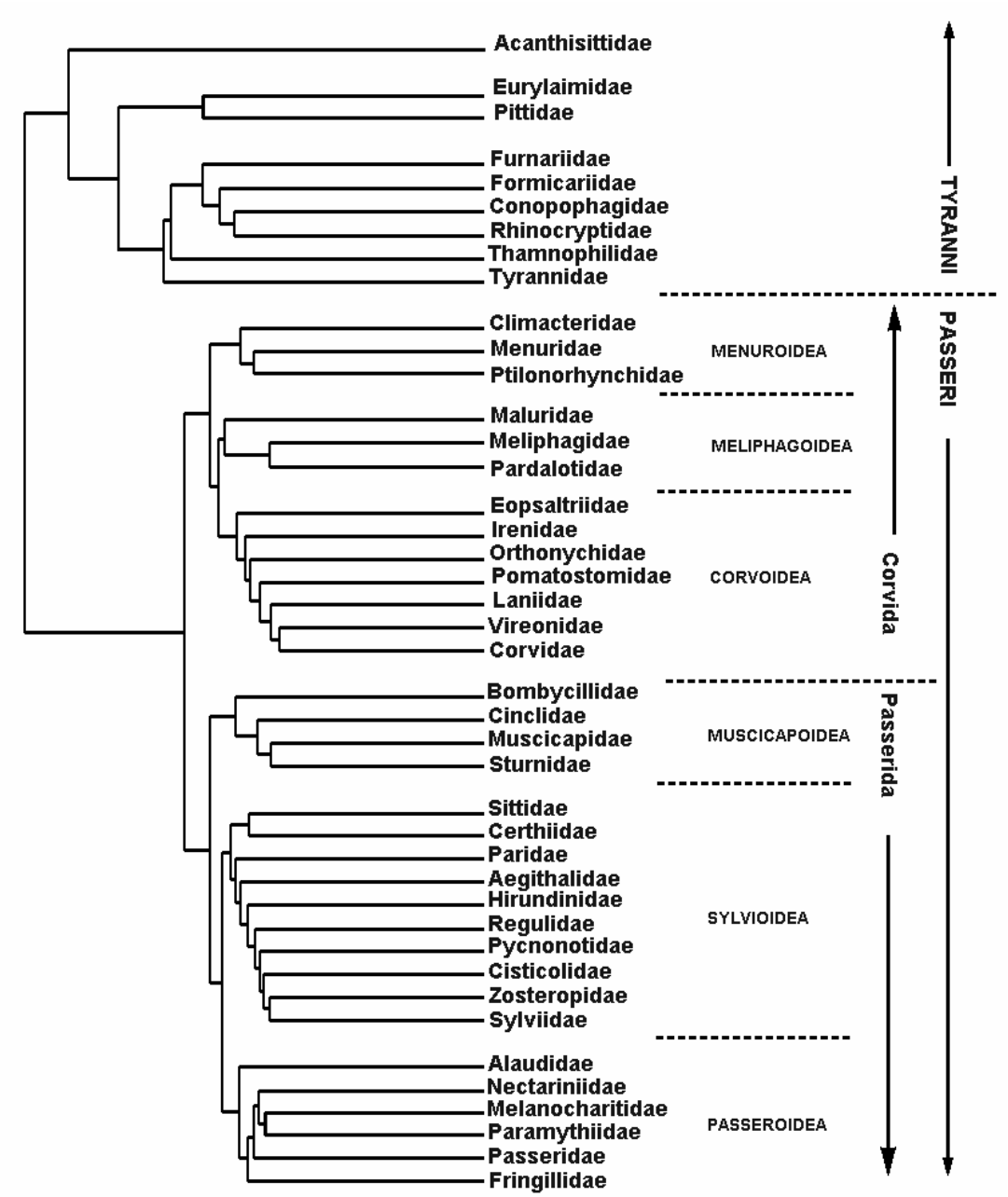


Fig. 1 Phylogenetic relationships of passerine families and their higher-level systematic classifications based on the DNA-DNA hybridization analyses of Sibley and Ahlquist (1990).

wrens (Acanthisittidae) has been established as the earliest branch within the Passeriformes and sister group to suboscines and oscines (Barker *et al.*, 2002; Ericson *et al.*, 2002a). The division of the oscines into the two sister taxa Corvida and Passerida, which had been hypothesised by Sibley and Ahlquist (1990), has been rejected later, as the Corvida appear to be paraphyletic (Barker *et al.*, 2002; Ericson *et al.*, 2002a, b). Additionally, conflicting phylogenetic hypotheses have been put forward for lower phylogenetic relationships, especially within the Passerida and their three superfamilies defined by Sibley and Ahlquist (1990): Muscicapoidea, Sylvioidea and Passeroidea (e.g. Barker *et al.*, 2004; Beresford *et al.*,

2005; Ericson *et al.*, 2003; Ericson and Johansson, 2003). For example, the phylogenetic position of the waxwings (Bombycillidae) at the basis of the Muscicapoidea has been questioned (e.g. Barker *et al.*, 2002; Ericson and Johansson, 2003). Within the Passeroidea, monophyly of Sibley and Ahlquist's (1990) Passeridae has been challenged repeatedly (Groth, 1998; Van der Meij *et al.*, 2005). The whole group of the Sylvioidea has been doubted, especially regarding the phylogenetic position of the kinglets (Regulidae), the clade consisting of treecreepers/wrens/nuthatches (Certhiidae and Sittidae), and the monophyly of the family Sylviidae (e.g. Barker *et al.*, 2002; Barker *et al.*, 2004; Ericson and Johansson, 2003). An additional point of concern has been the phylogenetic position of the two rockfowl species (Picathartidae, genus *Picathartes*), which for a long time has remained enigmatic.

While recent studies on the systematics of the whole order Passeriformes typically differ in their taxonomic sampling (at most, 173 passerine taxa were included in Beresford *et al*. (2005)), they generally rely on one or only a few nuclear genes as phylogenetic markers. Genes most commonly used have been: the single-copy recombination activating genes RAG-1 (Barker *et al.*, 2002; Barker *et al.*, 2004; Beresford *et al.*, 2005; Ericson and Johansson, 2003; Irestedt *et al.*, 2002; Irestedt *et al.*, 2001), and RAG-2 (Barker *et al.*, 2004; Beresford *et al.*, 2005), as well as the proto-oncogene c-*myc* (e.g. Ericson and Johansson, 2003; Ericson *et al.*, 2000; Irestedt *et al.*, 2002; Irestedt *et al.*, 2001), which encodes for a protein transcription factor, and myoglobin (Ericson and Johansson, 2003; Irestedt *et al.*, 2002). Although the advantages of combining different unlinked genes are well established (e.g. Moore, 1995), only a few studies have combined more than two molecular markers (e.g. Ericson *et al.*, 2002b), or added the mitochondrial marker cytochrome *b* (e.g. Ericson and Johansson, 2003). This latter gene showed evidence of saturation and has been found to be too variable for higher-level passerine systematics (e.g. Chikuni *et al.*, 1996; Edwards *et al.*, 1991; Edwards and Wilson, 1990). Despite all of these studies, many aspects of the phylogeny within the Passeriformes still remain unresolved, and often new ambiguities arise when additional species are included (Beresford *et al.*, 2005; Fuchs *et al.*, 2006).

Thus, in order to advance the clarification of passerine phylogenies, new molecular markers are needed. Therefore, I used one new nuclear gene (ZENK) and several chicken repeat 1 (CR1) retrotransposons as phylogenetic markers in passerine birds in addition to three nuclear protein-coding genes already established as phylogenetic markers (RAG-1, RAG-2, and c-*myc*).

1.2 ZENK and CR1 as new phylogenetic molecular markers

ZENK is a single-copy nuclear transcription factor expressed in the song system of birds and well-studied in the context of neurobiology (reviewed by Clayton, 1997; Ribeiro and Mello, 2000). ZENK, which is encoded by an immediate-early gene (IEG), is an acronym derived from the first character in the names of already described mammalian IEG homologs, i.e., the rodent **Z**if268 (Christy *et al.*, 1988), **E**gr-1 (Sukhatme *et al.*, 1988), the human **N**gfi-a (Milbrandt, 1987), and the rodent **K**rox-24 (Lanfear *et al.*, 1991), all of which share conserved sequence elements (Long and Salbaum, 1998). Expression of ZENK plays an important role in neuronal growth regarding learning and memory formation (reviewed by Ribeiro and Mello, 2000; Stork and Welzl, 1999; Tischmeyer and Grimm, 1999) and has been used as a marker of neuro-activity during song learning and production (reviewed by Ball and Gentner, 1998; Clayton, 1997). No evidence for selection pressure acting differentially on ZENK across diverse avian lineages has been found, despite the functional role of ZENK in avian physiology (Chubb, 2002; cited in Chubb, 2004a). Although it has been known since 1998 that this single-copy gene and parts of its 3' untranslated region (UTR) are highly conserved (Long and Salbaum, 1998), its use as a molecular marker in avian phylogenetics has been very limited so far. In a recent study, Chubb (2004a, b) demonstrated the usefulness of ZENK for higher level phylogeny in neognath birds as well as for the avian taxa Apodiformes (hummingbirds and swifts) and Passeriformes. The author provided evidence that ZENK is a powerful molecular marker with an estimated resolution for deep divergences within orders ranging roughly from 60 to 10 Mya. This analysis included only 18 passerine taxa and therefore obviously did not deliver a detailed phylogenetic hypothesis for the by far largest avian taxon.

The second newly established markers I used, chicken repeat 1 (CR1) elements (Stumph *et al.*, 1981), are repetitive DNA sequences. Interspersed repeats are very ubiquitous in the mammalian genome (40-50 %, IHGSC 2001; MGSC 2002), but with 9 % are comparably rare in the chicken genome (ICGSC 2004). A large number of these repetitive sequences are associated with mobile elements that can move from a parent locus to a target locus on the DNA level via DNA or RNA intermediates (Shedlock and Okada, 2000); this relocation process is called transposition. Classification and characteristics of mobile elements are shown in Figure 2. To differentiate between the two intermediate forms and to emphasise the reverse flow of genetic information, RNA mediated transposition is termed retrotransposition. Retrotransposons can be divided into a viral (containing retroviruses, long

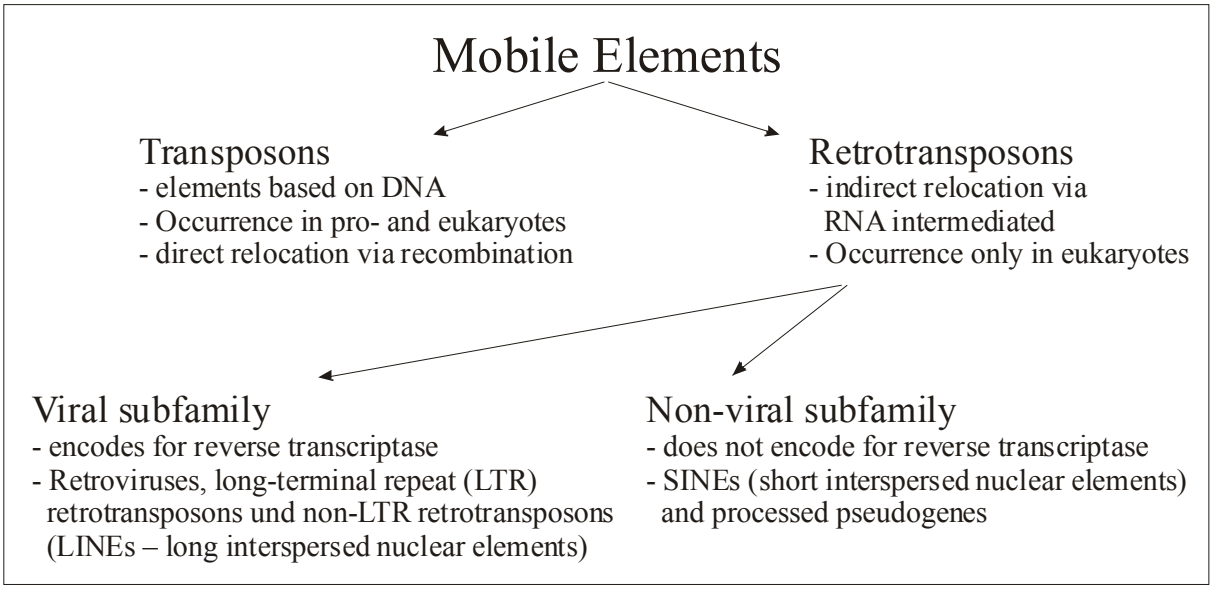


Fig. 2 Classifications and characteristics of different kinds of mobile elements. Classifications following definitions of Shedlock and Okada (2000).

terminal repeat (LTR) retrotransposons and non-LTR retrotransposons), and a nonviral superfamily (containing processed pseudogenes and short interspersed nuclear elements (SINEs, Shedlock and Okada, 2000)). Retrotransposons are widely dispersed throughout the genome and no process is known which could remove an inserted element from a locus. Thus, the prospect of using retrotransposons as phylogenetic markers seems very promising, because the presence of an element at a specific locus in two related species can be interpreted as a virtually homoplasy-free synapomorphy (Shedlock and Okada, 2000). The well-established use of SINE insertions as reliable apomorphic characters for phylogenetic inference in non-avian taxa (e.g., Huchon *et al.*, 2002; Lum *et al.*, 2000; Nikaido *et al.*, 2001; Nikaido *et al.*, 1999; Sasaki *et al.*, 2004; Schmitz *et al.*, 2001; Shedlock *et al.*, 2000; Shimamura *et al.*, 1997) was recently applied to CR1 insertions. For example, one single insertion in the lactate dehydrogenase B gene was found to support the monophyly of the Coscoroba/Cape Barren goose clade within the Anseriformes (St. John *et al.*, 2005), and a CR1 subfamily was analysed to resolve the phylogeny of penguins (Watanabe *et al.*, 2006). CR1 retrotransposon insertions constitute the largest amount of these mobile elements with more than 80 % (up to 200,000 copies in the chicken genome) and are the most important non-LTR retrotransposon in birds (ICGSC 2004). Figure 3 shows a schematic structure of a complete CR1 element. It possesses an 8 bp direct repeat at the 3'-end (typically [CATTCTRT] $[GATTCTRT]_{1\text{-}3}$ with some known variations), which can easily be detected (Silva and Burch, 1989). Two closely spaced open reading frames (ORF) have been found in the first complete consensus CR1 sequence (Burch *et al.*, 1993; Haas *et al.*, 1997). The first

ORF (ORF1) follows a 5'-untranslated region (UTR), which probably acts as a promoter (Haas *et al.*, 2001) and codes either for a zinc finger motif (Kajikawa *et al.*, 1997) or a nucleic acid binding protein (Haas *et al.*, 1997). The second ORF (ORF2) codes for an endonuclease and a reverse transcriptase (Haas *et al.*, 1997; Kajikawa *et al.*, 1997). A region of high sequence conservation is located near the end of the reverse transcriptase, which has been suggested to act as transcriptional silencer (Chen *et al.*, 1991). Additionally, parts of the 3'-untranslated region of CR1 elements show high sequence conservation and may serve as a protein binding site for a nuclear protein of unknown identity (Sanzo *et al.*, 1984). Thus, CR1 elements meet the criteria, which have been put forward by Eickbush (1992), that define them as non-LTR retrotransposons (Burch *et al.*, 1993). Until recently, only one full-length (4.5 kb) CR1 element with both intact ORFs has been described (ICGSC 2004). The first study on the evolution of CR1 elements resulted in the description of at least six different subfamilies (A-F) (Vandergon and Reitman, 1994). This was later expanded to 11 complete CR1 source genes and subdivided into 22 subfamilies (ICGSC 2004). These results pointed to a hypothesised ancient origin of these elements (Vandergon and Reitman, 1994), and were confirmed and extended by finding CR1 elements in the genomes of other vertebrates (Chen *et al.*, 1991; Fantaccione *et al.*, 2004; Kajikawa *et al.*, 1997; Poulter *et al.*, 1999), while CR1-like elements even have been reported in several invertebrate species (Albalat *et al.*, 2003; Biedler and Tu, 2003; Drew and Brindley, 1997; Malik *et al.*, 1999). The vast majority of CR1 elements have severely truncated 5'-ends and have lost their retrotransposable ability (Silva and Burch, 1989; Stumph *et al.*, 1981). After the insertion of a retrotransposable element at a specific locus in the genome of a common ancestor and the loss of the retrotransposable function by truncation, sequence evolution should not be constrained by selective pressure. This constitutes the possibility of using retrotransposon sequences as neutral molecular markers, apart from the established method of presence/absence screening. To my knowledge, such an approach has not been performed so far in a phylogenetic study of any vertebrate group.

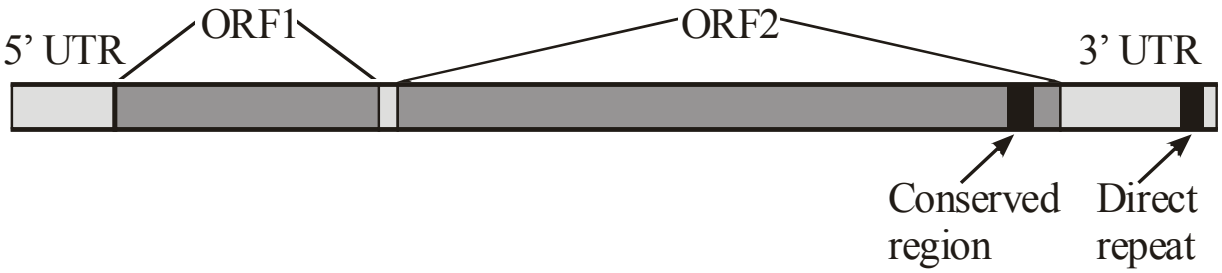


Fig. 3 Schematic structure of a complete chicken repeat 1 retrotransposon.

1.3 Aims of this study

The major aim of this study was to establish new molecular markers for avian systematics, apply them to the largest avian order (Passeriformes), and to provide new insights into passerine phylogenetic relationships. This complex and diverse taxon is well-studied and thus, provides useful information about proposed and conflicting phylogenetic hypotheses. For my dissertation research, I used three different approaches to contribute to the ongoing phylogenetic debate in the Passeriformes.

(1) I tested the recently introduced new molecular marker ZENK for its phylogenetic usefulness for passerine systematics in comparison to already established nuclear gene markers. The data set included representatives of as many passerine families as possible, i.e. 28 families and 40 subfamilies, with an emphasis on representatives of the Passerida. By using several different methods to create phylogenetic trees, I aimed at yielding the most robust phylogenetic results possible compared to existing phylogenetic hypotheses. A specific clade can be regarded as robust, if it is supported significantly and if different analyses generate the same topology. Therefore, I analysed data sets of single loci, as well as used a total evidence approach. Additionally, I investigated the phylogenetic utility of each marker by studying their levels of homoplasy and their contribution to the resolved nodes. I evaluated new or conflicting phylogenetic results by statistical tests.

(2) I have been the first to employ the clear-cut phylogenetic expressiveness of CR1 insertions as apomorphic characters in passerine systematics. I screened for specific CR1 loci in the raven *Corvus corax*. Two phylogenetic informative elements were detected in related taxa. I used the presence/absence pattern of these elements to help elucidate a special aspect of the phylogenetic puzzle, namely the position of the two African endemic rockfowl species *Picathartes oreas* and *Picathartes gymnocephalus* in the passerine tree. During this process, I found evidence that CR1 sequences contained a phylogenetic signal.

(3) The prospect of finding a phylogenetic signal in CR1 sequences provided the basis for my third approach. I detected and sequenced several CR1 elements isolated from Passeriformes in closely related species. I used these data to construct phylogenetic trees, compared, and analysed sequence composition and divergences. To appreciate the variability and divergences of CR1 sequences and to evaluate how meaningful the resulting phylogenetic trees were, I compared these to those calculated using sequences of established nuclear markers.

2 Summary of articles

2.1 Summary of article I:

SIMONE TREPLIN, ROMY SIEGERT, CHRISTOPH BLEIDORN, HAZELL SHOKELLU THOMPSON, ROGER FOTSO, AND RALPH TIEDEMANN.
Looking for the 'best' marker: songbird (Aves: Passeriformes) phylogeny based on sequence analyses of several unlinked nuclear loci.
Systematic Biology, submitted.

In this study I present a comprehensive phylogenetic analysis of a combination of established molecular markers (RAG-1, RAG-2, c-*myc*) and the recently introduced ZENK. The complete combined data set comprised 6,179 bp and included 80 taxa. I conducted phylogenetic analyses using maximum parsimony (MP, Farris *et al.*, 1970), maximum likelihood (ML, Felsenstein, 1981), and Bayesian inference (Huelsenbeck *et al.*, 2000; Larget and Simon, 1999; Mau and Newton, 1997; Mau *et al.*, 1999; Rannala and Yang, 1996). My analyses were performed using each gene separately and within a combined data set. I analysed the contribution of each gene on the phylogenetic tree yielded by the combined approach using partitioned Bremer support (PBS, Baker and DeSalle, 1997; Baker *et al.*, 2001; Baker *et al.*, 1998). This analysis evaluates the phylogenetic usefulness of the four genes. The ZENK trees exhibited by far the best resolution and showed the lowest amount of homoplasy compared to the other genes. My data indicate that this gene is – at least in passerines – suitable for inference even of ancient taxonomic splits, dating before the Cretaceous/Tertiary boundary.

The combined analysis yielded well-supported phylogenetic hypotheses for passerine phylogeny and, apart from corroborating recently proposed hypotheses on phylogenetic relationships within the Passeriformes, I provide evidence for several phylogenetic hypotheses: (1) The main passerine clades of suboscines and oscines are corroborated (2) just as the paraphyly of the Corvida. (3) Based on my study, I suggest a revision of the taxa Corvidae and Corvinae as vireos are closer related to crows, ravens, and allies. (4) I confirmed the subdivision of the Passerida into three superfamilies, Sylvioidea, Passeroidea, and Muscicapoidea, the first as a sister taxon to the two latter groups. (5) I found evidence for a strongly supported split within the Sylvioidea into two clades, one consisting of the tits

(Paridae) and the other comprising the bulbuls (Pycnonotidae), warblers, laughingthrushes, whitethroats, and allies (Timaliidae, *sensu* Alström *et al.*, 2006). (6) I suggest reflecting this split in a new classification of the Sylvioidea. (7) Additionally, my data point to a closer relationship between the Pycnonotidae and the Timaliidae than previous studies have indicated. (8) In my study, the Passeridae appear to be paraphyletic, because the finches (Fringillidae) are nested within the sparrows, wagtails, and pipits. (9) The monophyly of the weavers (Ploceinae) and the estrild finches (Estrildinae) as a separate, not yet described and named clade was strongly supported. (10) The sister taxon relationships of the dippers (Cinclidae) to the thrushes and flycatchers (Muscicapidae) was corroborated. (11) Finally, my data suggest a closer relationship of the waxwings (Bombycillidae) and the kinglets (Regulidae) to the wrens, tree-creepers (Certhiidae), and nuthatches (Sittidae).

The contributions of the different authors were as follows:

I performed the lab work for the c-*myc* data set, analysed the data, and wrote the manuscript. I established the methods and prepared lab work for R. Siegert, as well as I guided her during performing the lab work for the ZENK data set and the RAG-1 and RAG-2 sequences added to the data sets from GenBank. C. Bleidorn was involved in data analyses. Together with R. Tiedemann, he participated in the discussion of the results and the preparation of the manuscript. H. S. Thompson and R. Fotso provided the important samples of both *Picathartes* species.

2.2 Summary of article II:

SIMONE TREPLIN and RALPH TIEDEMANN.
Specific chicken repeat 1 (CR1) retrotransposon insertion suggests phylogenetic affinity of rockfowls (genus *Picathartes*) to crows and ravens (Corvidae).
Molecular Phylogenetics and Evolution, under review.

For this study I specifically screened for CR1 loci in Passeriformes and present two new CR1 loci found in the genome of the raven (*Corvus corax*). Sequences of these loci, named Cor1-CR1 and Cor2-CR1, are 372 bp and 283 bp in length, and belong to the 5' truncated CR1 elements. I used PCR to amplify these elements with specifically designed primers in several species closely related to the raven. The Cor1-CR1 locus was found additionally in representatives of the Corvinae (jays, crows, and allies), and thus corroborates monophyly of three tribes of the Corvinae, namely Corvini, Artamini, and Paradisaeini. The Cor2-CR1 locus could also be detected in orioles and two rockfowl species (genus *Picathartes*). The rockfowls are endemic to the West-African rainforest and consist of two species, the grey-necked picathartes (*Picathartes oreas*) and the white-necked picathartes (*Picathartes gymnocephalus*), which have long been regarded as avian curiosities (Thompson and Fotso, 1995). The phylogenetic position of these species within Passeriformes has been the object of extensive debate and for a long time has remained a puzzle, due to their unique combination of morphological traits. *Picathartes gymnocephalus* was originally described as a crow (*Corvus gymnocephalus*, TEMMINCK 1825) before being assigned to its own genus *Picathartes* (LESSON 1828). Rockfowls were alternately placed within babblers (Amadon, 1943; Delacour and Amadon, 1951), starlings (Lowe, 1938), corvids (Sclater, 1930) and thrushes (Amadon, 1943). Sibley and Ahlquist (1990) remained unsure about the phylogenetic position of *Picathartes* and *Chaetops* spp., the rockjumpers of South Africa and the closest relative to the rockfowls, and granted them a separate parvorder with the status of *incertae sedis*, aside all other Passeri. *Chaetops* itself has usually been placed among babblers (McLachlan and Liversidge, 1978; Sclater, 1930; Sharpe, 1883) and thrushes (Swainson, 1832). Recent sequence-based studies found that the Picathartidae (*Picathartes* and *Chaetops*) make up the earliest branch of the Passerida (Barker *et al.*, 2004; Beresford *et al.*, 2005; Ericson and Johansson, 2003). Thus, my results may provoke further discussion about the phylogenetic relationships at the boundary between the 'Corvida' and the Passerida. Nevertheless, as the Cor2-CR1 locus constitutes a synapomorphy for the three tribes Corvini,

Artamini, and Paradisaeini, together with the Oriolini and the Picathartidae, my study provided new evidence for a closer relationship of these species. Additionally, I showed that not only the absence/presence pattern of a CR1-insertion, but also the CR1-sequences themselves contain phylogenetic information.

The contributions of the different authors were as follows:

I performed all the lab work, analysed the data and wrote the manuscript. R. Tiedemann discussed the data with me and took part in the preparation of the manuscript.

2.3 Summary of article III:

SIMONE TREPLIN and RALPH TIEDEMANN.
Phylogenetic utility of chicken repeat 1 (CR1) retrotransposon sequences in passerine birds (Aves: Passeriformes).
Manuscript.

After I had discovered that CR1 sequences contained phylogenetic information (Article II), I wanted to investigate this issue in more detail. I screened genomes of three passerine species (the great tit, *Parus major*, the song thrush, *Turdus philomelos*, and the European pied flycatcher, *Ficedula hypoleuca*) for chicken repeat 1 (CR1) elements. I isolated seven CR1 loci varying in length, was able to design locus specific primers, and amplified those loci in several species other than the source organism. Additionally, I found a CR1 locus in GenBank that previously had been overlooked, by doing a blast search with my own CR1 sequences. I found this locus in Darwin's finches in reverse complement direction adjacent to a nuclear pseudogene of the mitochondrial cytochrome *b* gene (Sato *et al.*, 2001). I developed new primers for this locus, named Darfin-CR1, because the originally described ones (for the complete sequence including the pseudogene (Sato *et al.*, 2001)) failed to yield PCR products in species other than finches, and I was able to amplify this locus in all families of Passeriformes. Each locus was evaluated with regard to sequence characteristics and saturation effects, and was phylogenetically analysed using the Bayesian approach and maximum parsimony. My specific CR1 loci were found in the same species of (1) Muscicapoidea and (2) Sylvioidea (10 and 21 species, respectively; see Table 2 in Article III). I combined my CR1 loci and the Darfin-CR1 to two data sets named Mus-CR1 and Syl-CR1, both 742 bp in length. I performed phylogenetic analyses for each locus separately and for the two combined data sets. I compared distances of CR1 alignments to those of the established nuclear markers RAG-1 and ZENK and found not only evidence for a high variability in CR1 elements, but additionally for a correlated substitution rate of CR1 sequences and nuclear genes in most cases. I did not find evidence for saturation effects. To investigate the phylogenetic contents of each data set I conducted a likelihood-mapping which is based on the analysis of quartet puzzling (Strimmer and von Haeseler, 1997). This analysis indicated a higher resolution of the phylogenetic tree using the Mus-CR1 data. While the Syl-CR1 tree suffered from unresolved and non-supported clades above the genus level, the Mus-CR1 tree was fully resolved. Both trees were not fully congruent with previous hypotheses. My

analyses pointed to a better resolution of larger data sets (i.e. more loci/longer sequences and further taxa included). Nevertheless, I was able to provide evidence for the phylogenetic utility of CR1 retrotransposon sequences with this third study. It offers the opportunity to use sequences developed for classical presence/absence retrotransposon studies, which have turned out to be unsuitable for this approach, nevertheless as phylogenetic markers.

The contributions of the different authors were as follows:

I performed all the lab work, analysed the data and wrote the manuscript. R. Tiedemann discussed the data with me and took part in the preparation of the manuscript.

3 Discussion

3.1 Utility of new molecular markers for Passeriformes systematics

3.1.1 ZENK

The phylogenetic utility of the immediate-early gene ZENK and its homologs in mammals and zebrafish was indicated for the first time by Long and Salbaum (1998). The usefulness of ZENK and parts of the highly conserved 3'UTR for avian systematics was demonstrated in a higher-level phylogenetic study of neognath birds (Chubb, 2004a). This was additionally investigated and corroborated within the avian orders Apodiformes (hummingbirds and swifts) and Passeriformes (Chubb, 2004b). Whereas these previous studies only included 17 and 18 taxa, respectively, my study, comprising 80 taxa, is the first comprehensive analysis of passerine systematics using ZENK.

My analyses yielded fully resolved relationships among the three passerine families Muscicapoidea, Passeroidea and Sylvioidea, unlike the unresolved phylogenetic tree of Chubb (2004b). Both MP and Bayesian values significantly supported monophyly of these clades (Fig. 1, Article I). Comparing both the MP and the Bayesian phylogenetic trees of ZENK, I observed only few inconsistencies, mainly among passerine families (Fig. 1, Article I). A large proportion of clades in the passerine ZENK tree was fully resolved in my analyses, and only a few basal relationships within the Sylvioidea and the Muscicapoidea remained unresolved. Although Chubb (2004a, b) has already demonstrated the value of ZENK as a molecular marker, it can be evaluated even better when compared to other genes established for passerine systematics.

My single gene analyses illustrated the individual power of each gene to resolve phylogenetic relationships of Passeriformes. Such approaches have been applied rarely so far, as only RAG-1 and c-*myc* have been evaluated separately in a study on suboscine systematics (Irestedt *et al.*, 2001). RAG-1 was supposed to have great potential in resolving ancient avian divergences, but failed in fast evolved lineages (Groth and Barrowclough, 1999; Irestedt *et al.*, 2001). RAG-2 has been used only in combination with RAG-1 so far (Barker *et al.*, 2004; Beresford *et al.*, 2005). In my analyses, the single-locus phylogenetic trees of RAG-2 and c-*myc* suffered from a high degree of unresolved nodes. I corroborated the usefulness of RAG-1 to resolve uncertain phylogenetic relationships. The values of the partitioned Bremer support (PBS) indicated that RAG-1 had contributed to most of the nodes of the maximum parsimony

strict consensus tree (supplementary data, Article I). Nevertheless, it was outperformed by ZENK, because the ZENK trees exhibited by far the best resolution of all genes analysed. The phylogenetic tree based on ZENK contained the largest number of resolved nodes and of nodes that were congruent with the phylogenetic tree of the combined data set (40, compared to 6-18 for the other three genes, Table 3, Article I). The PBS values, however, indicated that ZENK did not dominate the combined data set. In the ZENK data set, observed levels of homoplasy were the lowest of all genes, which further adds to its superior ability to resolve passerine phylogenies (Table 3, Article I). The PBS values indicated only a slightly smaller contribution of ZENK to the combined data set compared to RAG-1. Resolving phylogenies within Passeroidea with ZENK consistently showed the highest PBS values among all nodes (supplementary data, Article I). This was reflected also in the phylogenetic tree of the single gene analysis of ZENK, where all nodes were resolved (Fig. 1, Article I).

According to Chubb (2004b), the highest power of the ZENK gene is in resolving lineages which diverged roughly 60 to 10 Mya ago. My data indicate that this gene is – at least in passerines – suitable for inference of even older taxonomic splits. The split into the suboscine taxa of Furnarioidea and Tyrannoidea is estimated to have occurred 61-65 Mya ago and into the suborders suboscines and oscines around 76 Mya ago (Barker *et al.*, 2004). These clades were resolved and strongly supported in my phylogenetic tree using ZENK. Thus, resolution of lineages, which originated before the Cretaceous/Tertiary boundary, is possible using the ZENK gene as well.

My study showed the advantages of using the ZENK gene and its 3' UTR region in a phylogenetic analysis of Passeriformes. Nevertheless, I would recommend performing a combined approach of different genes as it was apparent that the combined data set was superior to all single-locus analyses in resolving passerine phylogenies.

3.1.2 CR1 elements as apomorphic markers

Since the first demonstrations of short interspersed element (SINE) insertions providing robust phylogenetic signal (e.g. Okada, 1991), this method has been expanded to a powerful tool for recovering monophyletic clades (e.g. Cook and Tristem, 1997; Rokas and Holland, 2000; Shedlock *et al.*, 2000). Verneau *et al*. (1998) and Nikaido *et al*. (1999) applied this approach successfully to non-LTR retrotransposons (LINEs), and were followed by studies using LINE-1 (L1) element insertions as phylogenetic markers (Lutz *et al.*, 2003; Mathews *et al.*, 2003; Vincent *et al.*, 2003). Despite the high abundance of chicken repeat 1

(CR1) retrotransposons (ICGSC 2004), only two studies have performed phylogenetic analyses with these elements in birds, namely in Anseriformes and Sphenisciformes (St. John *et al.*, 2005; Watanabe *et al.*, 2006, respectively). The two CR1 elements I found in the raven, *Corvus corax*, appeared appropriate for inferring phylogenetic relationships (Article II).

The difficulties in my approach of analysing the absence/presence pattern of CR1 loci consisted in the truncated 5'-ends of the elements. Different opinions have been proposed whether CR1 elements create target site duplications: whereas Silva and Burch (1989) proposed that such duplications can always be found, Vandergon and Reitman (1994) limited this event to only some CR1 elements, and recently it was suggested never to occur (ICGSC 2004). Detection of such a duplication and hence, identification of the 5'-end was impossible in the Cor1- and Cor2-CR1 loci and thus, I was unable to perform a classical presence/absence screening as one primer was lying within the element and the other, i.e. the locus-specific primer, in the 3' flanking region. As PCR yielded single-locus products my strategy of '3'-flanked PCR' does work. According to Shedlock and Okada (2000), false negative results do not challenge the phylogenetic relationships of those species for which positive PCR amplifications have been obtained. As an independent control, I confirmed negative results by performing hybridisation experiments (Fig. 4, Article II). Additionally, I solved the problem of false positive signals, like the finding of the Cor2-like-CR1 element in the Bohemian waxwing, *Bombycilla garrulus*, and the white-throated dipper, *Cinclus cinclus*, by directly sequencing the PCR products. The differing 3'-flanking region unambiguously pointed to a different locus (Fig. 3, Article II).

To avoid these difficulties in future studies, two possibilities are obvious: (1) Similar to the study of St. John *et al*. (2005), one could use CR1 elements, which have been inserted in introns of coding genes. This provides unambiguous ends of the elements and facilitates primer-design in conserved regions of the gene of concern. As it appears rather unlikely to find such an intron in the genome of the taxon of interest, (2) screening a genomic library would possibly be more successful when concentrating on the development of longer clones, since this increases the likelihood of yielding sequences containing both ends of the elements. Nevertheless, I consider presence of my newly discovered CR1 loci an apomorphic character state, proving that these elements can be used to infer phylogenetic relationships within Passeriformes in general.

3.1.3 Sequences of CR1 elements

The gain of using retrotransposon insertions as noise-free apomorphic phylogenetic characters is often disproportionate to the effort one has to invest finding enough suitable elements. It is known that retrotransposon subfamilies have had different rates of transposition activity. An appropriate marker has to have been active specifically during the time of divergence of a clade in question (e.g. Kido *et al.*, 1991; Sasaki *et al.*, 2004; Shimamura *et al.*, 1997). Searching for such elements, one will inevitably find many apomorphic, but uninformative markers (with regard to the specific question), e.g. those that are found in all representatives of the investigated group. It is widely accepted that retrotransposons accumulate neutral substitutions after an insertion event, in particular after losing their retrotransposition ability, like CR1 elements, (Kido *et al.*, 1995; Webster *et al.*, 2006). Thus, I hypothesised that these sequences contain a phylogenetic signal. This offered the opportunity to use the retrotransposon sequences themselves as a phylogenetic marker.

I could successfully apply this approach in my study on the insertion pattern of two CR1 loci (Figs. 5 and 6, Article II). Furthermore, the eight CR1 loci I investigated in regard to their phylogenetic utility (Article III) obviously lost their retrotransposable ability as indicated by several conspicuous indels in the region of ORF1. These elements did not completely match a sequence of reverse transcriptase. Thus, random mutation must have caused the high variability in the CR1 elements among the species studied. I assessed the variability of CR1 sequences by comparing them to the two genes ZENK and RAG-1. Substitution rates of CR1 sequences were up to 3.2 times higher than those of ZENK were, and variability in the two marker systems was correlated significantly in most cases (Fig. 2, Article III). It usually is assumed that markers with high variability are saturated due to multiple substitutions. This has been shown for the mitochondrial cytochrome *b* gene, which consequently was used less frequently to resolve higher-level phylogenies in Passeriformes (Chikuni *et al.*, 1996; Edwards *et al.*, 1991; Edwards and Wilson, 1990). However, I did not find any indications of saturation in my CR1 loci when comparing transitions (ti) and transversions (tv) to total sequence divergences, and neither did the ti/tv ratio point to multiple substitutions (Fig. 1 and Table 3, Article III). My data further indicated a very low level of homoplasy in the CR1 sequences (Table 3, Article III). These sequence characteristics indicated a powerful phylogenetic signal. The method of likelihood-mapping visualised the phylogenetic signal and corroborated my hypothesis with different results for my two combined CR1 data sets. According to these findings, the Mus-CR1 data set is superior to the Syl-CR1 data set in resolving phylogenies. Even though likelihood-mapping does not always produce fully

reliable results (Nieselt-Struwe and von Haeseler, 2001), the Mus-CR1 tree and the less resolved phylogenetic tree of the Syl-CR1 data set corroborated this analysis (Figs. 4 and 5, Article III).

Despite the evidence for CR1 sequences containing useful phylogenetic information, the phylogenetic trees were not fully congruent with recent hypotheses about passerine systematics and showed some relationships, which are supposed to be unlikely. Possibly, these particular data sets were too small, especially for resolving the taxon of Sylvioidea, which has been shown to be difficult to elucidate (Alström *et al.*, 2006; Jønsson and Fjeldså, 2006; Sheldon and Gill, 1996). As the rather short sequences of the single CR1 loci failed to produce unambiguous trees, including sequences of additional CR1 loci presumably would increase the phylogenetic signal. As there is such a high number of CR1 elements in the genomes of birds, generating larger data sets (i.e. more loci/longer sequences and including additional taxa) than the ones in my study, could definitely contribute to the ongoing debate on passerine phylogenies. Specific screens for retrotransposons as sequence markers may be useful for studies, where previous marker systems have been less successful.

3.2 Phylogenetic relationships within Passeriformes inferred from new markers

3.2.1 Suboscines

My so far partial taxon sampling of suboscines and non-Passerida oscines allows only an incomplete phylogenetic inference for these groups and, thus, will be discussed only briefly. Sibley and Ahlquist (1990) found a split of the New-World suboscines in the three clades Tyrannida, Furnariida, and typical antbirds (Thamnophilidae). I did not find support for this partition, instead I support the integration of the typical antbirds into the Furnariida, as well as monophyly of the ovenbirds and woodcreepers (Furnariidae) and their sister group relationship with the ground antbirds (Formicariidae) (Chesser, 2004; Irestedt *et al.*, 2002; Irestedt *et al.*, 2001; Article I).

3.2.2 'Corvida'

This taxon was the most surprising new classification proposed by Sibley and Ahlquist (1990), because it comprised several species with different morphological traits and geographical distribution. Nevertheless, it was accepted at first (Lovette and Bermingham, 2000). Later, however, its monophyly has been doubted by several authors and the Corvida have generally been rendered paraphyletic (Barker *et al.*, 2002; Ericson *et al.*, 2002a, b). This was confirmed by my study, as the honeyeaters (Meliphagidae) (originally included in 'Corvida' by Sibley and Ahlquist (1990)) are identified as a sister taxon to all other oscines (Figs. 1-4, Article I). I found that the orioles (Oriolini) are not as closely related to the ravens, crows, jays, and allies (Corvini), as had been hypothesised by Sibley and Ahlquist (1990) (Article I and II) and therefore I am challenging their taxon Corvinae, consisting of the tribes Corvini, Artamini (currawongs), Paradisaeini (birds of paradise), and Oriolini. In Article I, I corroborated the sister taxon relationship between birds of paradise and corvids previously hypothesised (Cracraft and Feinstein, 2000; Frith and Beehler, 1998; Helmbychowski and Cracraft, 1993; Nunn and Cracraft, 1996). In addition, I propose that the taxon Corvidae (*sensu* Sibley and Ahlquist, 1990) needs to be revised, because vireos (Vireonidae) are apparently closely related to the corvids and might even be nested within the Corvidae (Article I). Thus, phylogenetic relationships within the Corvidae remain unresolved and need further investigation, preferably with a more complete taxon sampling.

3.2.3 Picathartidae

The different historical classifications of the genus *Picathartes* (see 2.2) illustrate the difficulties in resolving its phylogenetic relationships. In this regard, the results of my CR1 insertion analyses (Article II) are at odds with my study based on sequences (Article I) and several other studies (Barker *et al.*, 2004; Beresford *et al.*, 2005; Ericson and Johansson, 2003). Sibley and Ahlquist (1990) tentatively concluded that *Picathartes* should have affinities to Corvida, as corroborated by my Cor2-CR1 insertion (Article II), but they conveyed their uncertainty, coupled with ambiguous morphological data, by placing the taxon Picathartidae (*Picathartes gymnocephalus*, *P. oreas*, *Chaetops frenatus*, *C. aurantius*) beside Corvida and Passerida with the status of *incertae sedis*. Ericson and Johansson (2003) proposed *Picathartes* and *Chaetops* being basal to the Passerida. They classified them as Passerida because they all share a 3 bp-insertion in the sequence of c-*myc*, a character considered apomorphic for the Passerida. Beresford *et al*. (2005) and Barker *et al*. (2004)

challenged this, by proposing the Petroicidae as the second branch in the Passerida (branching off after the Picathartidae), because the Petroicidae lack this insertion (only available representative *Eopsaltria australis* (Ericson *et al*., 2002b)). Recently, Fuchs *et al*. (2006) and Jønsson and Fjeldså (2006) discussed the difficulties in recovering a robust phylogenetic hypothesis at the boundary between 'Corvida' and Passerida using sequence data. Regarding these ambiguous data, both morphological and molecular, and the clear-cut character state of the Cor2-CR1 locus, my analyses suggest a closer relationship of the Picathartidae to the Corvidae.

3.2.4 Passerida

My analyses strongly corroborated the partition of Passerida into three superfamilies Passeroidea, Muscicapoidea, and Sylvioidea (originally defined by Sibley and Ahlquist (1990)), however, with slight modifications (Article I).

Passeroidea.–The major differences in phylogenetic relationships within the Passeroidea compared to those established by Sibley and Ahlquist (1990) was in the inclusion of fairy-bluebirds and leafbirds (Irenidae), (which had been classified as 'Corvida' by Sibley and Ahlquist (1990)) and the exclusion of the larks (Alaudidae) (Article I). Apart from these fundamentally new classifications, my study also pointed to a revision at lower phylogenetic levels. Sibley and Ahlquist's (1990) family Passeridae should not be maintained, because their family Fringillidae is embedded in parts of the Passeridae (Article I). According to Sibley and Ahlquist (1990), this taxon consists of five subfamilies, namely (1) sparrows (Passerinae), (2) wagtails and pipits (Motacillinae), (3) accentors (Prunellinae), (4) weavers (Ploceinae), and (5) estrildine finches (Estrildinae). I found strong support for a split of the Passeridae into two clades, one consisting of sparrows, wagtails, and pipits (subfamilies 1 and 2) and the other consisting of weavers and estrildine finches (subfamilies 4 and 5) (Article I). This relationship has been postulated previously, albeit with high uncertainty (Groth, 1998) and was recently corroborated (Van der Meij *et al*., 2005). My analyses significantly supported the monophyletic clade of weavers and estrildine finches, and I found support for the position of the whydahs (Viduini) as the basal branch of the estrildine finches (Figs. 1-4, Article I), a placement considered controversial (Groth, 1998; Sibley and Ahlquist, 1990). Due to incongruence among different analysis methods, my results so far are ambiguous regarding the phylogenetic position of the accentors (Article I). Their position as the earliest branch of the Passeridae and the Fringillidae has been suggested previously (Barker *et al.*,

2004; Beresford *et al.*, 2005; Ericson and Johansson, 2003). In contrast, a closer relationship to sparrows was supported by the ZENK data set and the MP bootstrap analysis of the combined data set (Figs. 1 and 2, Article I). My data definitely rejects the hypothesis of accentors being closer related to weavers and estrildine finches, which has been found in the supertree of Jønsson and Fjeldså (2006).

Muscicapoidea.–My studies strongly corroborated recent findings about the phylogeny of the Muscicapoidea. If one accepts the exclusion of the waxwings (Bombycillidae) from this taxon (as discussed below), higher-level relationships seem to consolidate with the starlings and mockingbirds (Sturnidae) as the earliest branch. In particular, I was able to validate the position of the dippers (Cinclidae) as a sister taxon to the Muscicapidae for the first time with significant MP support (Figs. 2 and 4, Article I). The split of the Muscicapidae into the two clades of thrushes (Turdinae) and the chat (Saxicolini)/flycatcher (Muscicapini) assemblage (Muscicapinae) was congruent with many other studies (e.g. Barker *et al.*, 2004; Beresford *et al.*, 2005; Cibois and Cracraft, 2004; Jønsson and Fjeldså, 2006). My data confirmed the monophyly of the chats and flycatcher, with the modification, that the European pied flycatcher *Ficedula hypoleuca* should be included into the chats (Article I and III). Originally, it had been classified as a member of the Muscicapini (Sibley and Monroe, 1990).

The waxwings recently have been referred to as a 'problem clade', which 'moves around' in the phylogenetic trees (Jønsson and Fjeldså, 2006). They have either not been resolved at all (Ericson and Johansson, 2003; Fuchs *et al*., 2006), associated with the tits (Paridae) as the deepest branch within the Sylvioidea (Barker *et al*., 2002), or have been placed basally within the Muscicapoidea (Barker *et al.*, 2004; Beresford *et al.*, 2005; Voelker and Spellman, 2004). Barker *et al*. (2004) showed an affinity of the waxwings to the kinglets (Regulidae), however with only little support. The kinglets themselves, classified as a member of the Sylvioidea by Sibley and Ahlquist (1990), were recently called another 'lost lineage' in the passerine tree (Jønsson and Fjeldså, 2006). I found additional evidence for the waxwings and the kinglets being closely related, possibly as sister taxa (Article I and III). A closer relationship of the waxwings and the clade of wrens, tree-creepers, and nuthatches (Certhiidae and Sittidae) has been adumbrated with these groups as deepest splits in the Muscicapoidea (Jønsson and Fjeldså, 2006). Wrens, tree-creepers, and nuthatches had been placed in the Sylvioidea by Sibley and Ahlquist (1990) and meanwhile, the closer relationship to the Muscicapoidea has been confirmed (Barker *et al.*, 2002; Barker *et al.*, 2004; Beresford *et al.*, 2005). In my study, I cautiously suggest the existence of a clade consisting of waxwings,

kinglets and the wrens/tree-creepers/nuthatches assemblage, but this hypothesis awaits further detailed investigations (Article I).

Sylvioidea.–Phylogenetic relationships within the second largest group of oscine birds, the Sylvioidea (*sensu* Sibley and Monroe, 1990) have been difficult to elucidate (Alström *et al*., 2006; Jønsson and Fjeldså, 2006). For example, the exact phylogenetic position of tits (Paridae) has frequently not been resolved, even in recent studies, and an exclusion from the Sylvioidea has been proposed (e.g. Alström *et al.*, 2006). Alström *et al*. (2006) suggested to apply the name 'Sylvioidea' to a clade without tits. My studies provided strong evidence for a robust tit-clade as the sister taxon to the Sylvioidea (Article I and III). If the denomination of 'Sylvioidea' should be retained, it would require a new name for this sister clade, and I suggest to assign the name 'Paroidea' to it, comprising the tits and relatives. Although Linnean categories (like superfamilies) are not based on absolute criteria, this new classification might ease further discussion on their respective phylogenetic relationships.

My results were ambiguous concerning the phylogenetic position of the larks (Alaudidae). When applying different analysis methods I found either (1) the larks together with the swallows (Hirundinidae) embedded in the Sylvioidea, or (2) the larks as the earliest branch of the Sylvioidea (Article I and III). As previous authors have proposed the second hypothesis I also assume it to be more likely (Alström *et al.*, 2006; Barker *et al.*, 2004; Beresford *et al.*, 2005; Ericson and Johansson, 2003; Fuchs *et al.*, 2006).

In recent studies, the leaf-warblers (Acrocephalinae) have appeared to be a polyphyletic group (e.g. Alström *et al*., 2006; Sefc *et al*., 2003). I strongly confirmed this by finding a *Phylloscopus*-warbler clade (Article I and III). Additionally, the leaf-warblers should be excluded from their original classification in Sylviidae (*sensu* Sibley and Ahlquist, 1990) (Article I and III). In fact they recently have been granted their own family-status (Acrocephalidae, Alström *et al.*, 2006). My data failed to unambiguously resolve the phylogenetic position of the common grashopper-warbler *Locustella naevia*. It forms a monophylum with the *Acrocephalus* sp./*Hippolais icterina* clade or constitutes a basal branch within the Sylvioidea depending on the analysis method (Figs. 2, 4, and 1, 3, Article I, respectively). Haffer (1991) suggested a close relationship of *Locustella*, *Acrocephalus*, and *Hippolais*, but meanwhile, this relationship has been questioned (Helbig and Seibold, 1999). Thus, the phylogenetic position of *Locustella* requires further investigation.

My study is the first to yield a highly resolved and strongly supported clade consisting of the bulbuls (Pycnonotidae) and the babblers, white-eyes, laughingthrushes, and allies

(Timaliidae) (Figs. 2 - 4, Article I). This relationship has been found before but until now has lacked statistical support (Barker *et al.*, 2002; Barker *et al.*, 2004; Beresford *et al.*, 2005). The newly defined family of the Timaliidae takes into account the non-monophyly of Sibley and Ahlquist's (1990) Sylviidae and the closer relationships of the white-eyes (Zosteropidae) to the babblers (Timaliini), laughingthrushes (Garrulacinae), and allies (Alström *et al.*, 2006). My data pointed to a sister taxon relationship between laughingthrushes and white-eyes and strongly corroborated the revision of the Sylviidae, with new evidence apart from the exclusion of the leaf-warblers (Article I and III).

Despite my comprehensive analyses, the clade of Sylvioidea could not be fully resolved. The short branch lengths and internodes in my phylogenetic trees (Figs. 1 - 4, Article I) and the fact that this group had previously been given the status of the least resolved group in the passerine supertree based on a metaanalysis of 99 studies (Jønsson and Fjeldså, 2006), point to a particularly rapid speciation and radiation of this group. Thus, future studies on the phylogenetic relationships within the Passeriformes should consider especially these species as a major subject of investigation.

3.3 Conclusion

My phylogenetic approaches using different new molecular markers further advance the ongoing debate about phylogenetic relationships of the Passeriformes. I present a revised phylogenetic tree of major passerine groups inferred from my studies in Figure 4. My comprehensive sequence analyses (Article I and III) and the study using CR1 insertions as apomorphic characters (Article II) have shown that these promising markers can contribute to phylogenetic studies of the Passeriformes. I was able to settle several controversial issues in passerine phylogenies. Furthermore, these markers may be applied to the molecular systematic of birds in general. Future studies should include an even more extensive taxon sampling to clarify the last remaining uncertainties.

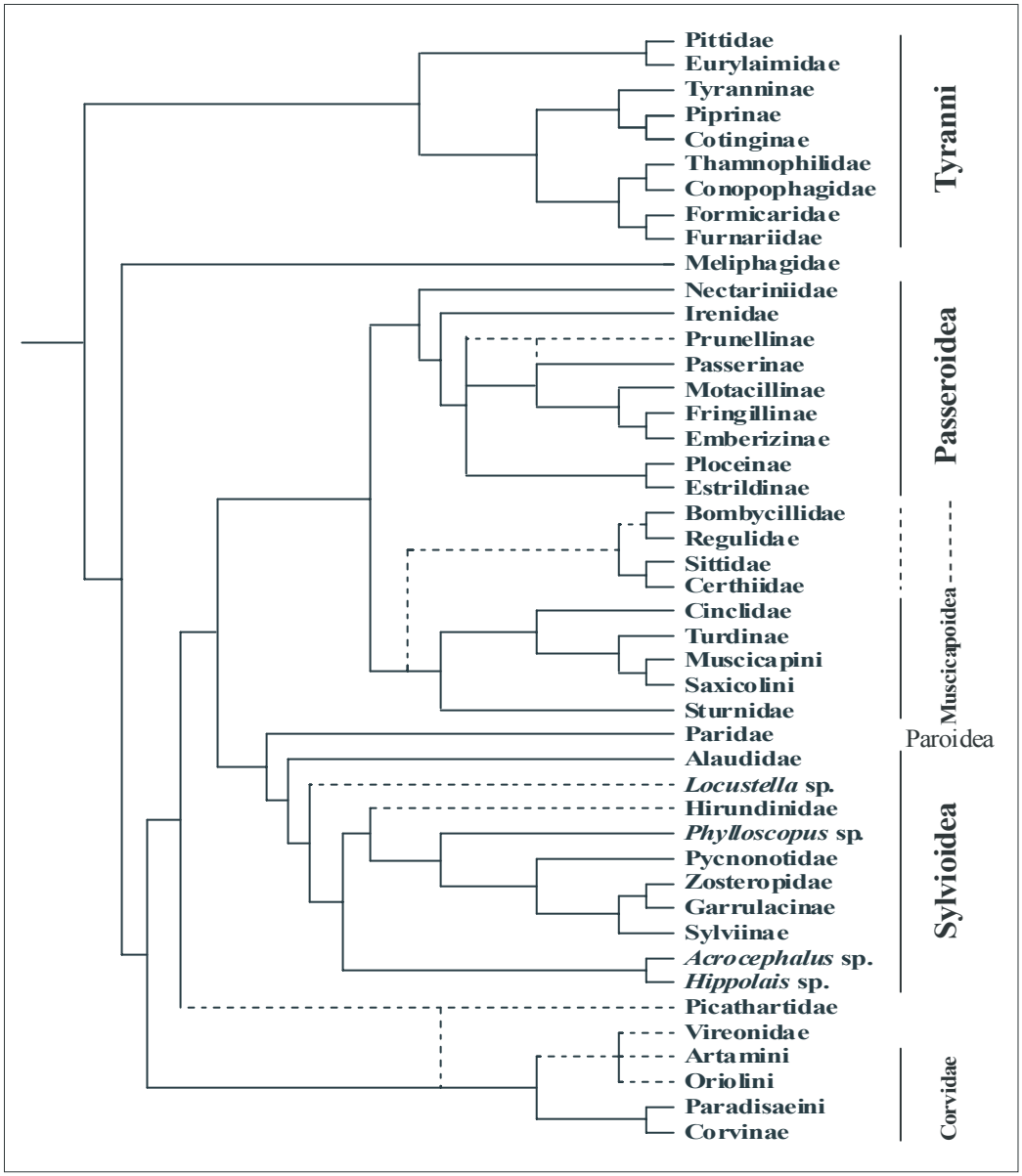


Fig. 4 Revised phylogenetic tree of the major passerine groups inferred from my results. The dashed branches indicate remaining uncertain phylogenetic relationships.

4 Abstract

The aim of this study was to provide deeper insights in passerine phylogenetic relationships using new molecular markers. The monophyly of the largest avian order Passeriformes (~59 % of all recent birds) and the division into its suborders suboscines and oscines are well established. Phylogenetic relationships within the group have been extremely puzzling, as most of the evolutionary lineages originated through rapid radiation. Numerous studies have hypothesised conflicting passerine phylogenies and have repeatedly stimulated further research with new markers. In the present study, I used three different approaches to contribute to the ongoing phylogenetic debate in Passeriformes. I investigated the recently introduced gene ZENK for its phylogenetic utility for passerine systematics in combination and comparison to three already established nuclear markers. My phylogenetic analyses of a comprehensive data set yielded highly resolved, consistent and strongly supported trees. I was able to show the high utility of ZENK for elucidating phylogenetic relationships within Passeriformes. For the second and third approach, I used chicken repeat 1 (CR1) retrotransposons as phylogenetic markers. I presented two specific CR1 insertions as apomorphic characters, whose presence/absence pattern significantly contributed to the resolution of a particular phylogenetic uncertainty, namely the position of the rockfowl species *Picathartes* spp. in the passerine tree. Based on my results, I suggest a closer relationship of these birds to crows, ravens, jays, and allies. For the third approach, I showed that CR1 sequences contain phylogenetic signal and investigated their applicability in more detail. In this context, I screened for CR1 elements in different passerine birds, used sequences of several loci to construct phylogenetic trees, and evaluated their reliability.

I was able to corroborate existing hypotheses and provide strong evidence for some new hypotheses, e.g. I suggest a revision of the taxa Corvidae and Corvinae as vireos are closer related to crows, ravens, and allies. The subdivision of the Passerida into three superfamilies, Sylvioidea, Passeroidea, and Muscicapoidea was strongly supported. I found evidence for a split within Sylvioidea into two clades, one consisting of tits and the other comprising warblers, bulbuls, laughingthrushes, whitethroats, and allies. Whereas Passeridae appear to be paraphyletic, monophyly of weavers and estrild finches as a separate clade was strongly supported. The sister taxon relationships of dippers and the thrushes/flycatcher/chat assemblage was corroborated and I suggest a closer relationship of waxwings and kinglets to wrens, tree-creepers, and nuthatches.

5 Abstract (German version)

Das Ziel dieser Arbeit war es, mittels neuer molekularer Marker zusätzliche Informationen über die phylogenetischen Verwandtschaftsverhältnisse der Sperlingsvögel (Passeriformes) zu erhalten. Die Monophylie der Passeriformes, der größten Vogelgruppe (~59 % aller lebenden Arten), sowie ihrer Unterteilung in Suboscines und Oscines sind gut belegt. Die phylogenetischen Verwandtschaftsverhältnisse innerhalb dieser Gruppen sind jedoch seit jeher sehr schwer zu entschlüsseln, da sich die meisten Linien durch eine schnelle Radiation entwickelten. Zahlreiche Studien haben verschiedene Hypothesen zur Phylogenie der Sperlingsvögel aufgestellt und damit die Suche nach neuen Markern initiiert. In meiner Untersuchung habe ich drei verschiedene Ansätze benutzt, um zur Klärung der Phylogenie beizutragen. Ich untersuchte das kürzlich als Marker eingeführte ZENK-Gen im Hinblick auf seinen Nutzen in der Systematik der Sperlingsvögel in Kombination und im Vergleich zu drei bereits etablierten nukleären Markern. Meine phylogenetischen Analysen eines umfassenden Datensatzes ergaben hoch aufgelöste, konsistente und stark unterstütze Stammbäume, so dass ich den hohen Nutzwert des ZENK-Gens für die Klärung phylogenetischer Verwandtschaftsverhältnisse der Passeriformes zeigen konnte. Für den zweiten und dritten Ansatz habe ich Chicken Repeat 1 (CR1) Retrotransposons als phylogenetische Marker benutzt. Anhand zweier spezifischer CR1 Insertionen als apomorphe Merkmale und deren Insertionsmuster in verschiedenen Sperlingsvögeln konnte ich die phylogenetische Position der afrikanischen Felshüpfer, *Picathartes* spp., klären. Aufgrund meiner Ergebnisse schließe ich auf eine engere Verwandtschaft der Felshüpfer zu den Rabenvögeln. Durch meinen dritten Ansatz konnte ich nachweisen, dass CR1-Sequenzen phylogenetische Informationen enthalten, und untersuchte detailliert deren Anwendung als Marker. Dafür habe ich in verschiedenen Sperlingsvögeln nach CR1 Elementen gesucht und mit einigen dieser Sequenzen Stammbäume berechnet, um die Verlässlichkeit der Marker zu überprüfen.

Durch meine Untersuchungen konnte ich existierende Hypothesen stützen und zusätzlich starke Hinweise auf neue Hypothesen finden. Beispielsweise schlage ich eine Revision der Taxa Corvidae und Corvinae vor, da Vireos eng mit den Rabenvögeln verwandt sind. Die Unterteilung der Passerida in die drei Unterfamilien Sylvioidea, Passeroidea und Muscicapoidea konnte deutlich bestätigt werden. Ich habe Hinweise auf eine Trennung der Sylvioidea in zwei taxonomische Gruppen erhalten, einer bestehend aus Meisen und Verwandten und der andere aus Grasmücken, Bülbüls, Häherlingen, Brillenvögeln und

Verwandten. Während die Passeridae paraphyletisch sind, wurde die Monophylie der Weber und Astrilden als ein eigenes Taxon unterstützt. Das Schwestergruppenverhältnis zwischen Wasseramseln und dem Drossel/Fliegenschnäpper/Schmätzer-Taxon wurde ebenfalls bestätigt. Außerdem habe ich Hinweise auf eine nähere Verwandtschaft zwischen Seidenschwänzen und Goldhähnchen zu Zaunkönigen, Baumläufern und Kleibern gefunden.

6 References

Albalat, R., Permanyer, J., Martinez-Mir, C., Gonzalez-Angulo, O., Gonzalez-Duarte, R., 2003. The first non-LTR retrotransposon characterised in the cephalochordate amphioxus, BfCR1, shows similarities to CR1-like elements. Cell. Mol. Life Sci. 60, 803-809.

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.

Amadon, D., 1943. The genera of starlings and their relationships. Am. Mus. Novit. 1247, 1-16.

Ames, P., 1971. The morphology of the syrinx in passerine birds. Bull. Peabody Mus. Nat. Hist. 37, 194.

Baker, R. H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46, 654-673.

Baker, R. H., Yu, X. B., DeSalle, R., 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol. 9, 427-436.

Baker, R. H., Wilkinson, G. S., DeSalle, R., 2001. Phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae). Syst. Biol. 50, 87-105.

Ball, G. F., Gentner, T. Q., 1998. They're playing our song: gene expression and birdsong perception. Neuron 21, 271-274.

Barker, F. K., Barrowclough, G. F., Groth, J. G., 2002. A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. Proc. R. Soc. Lond. B Biol. Sci. 269, 295-308.

Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., Cracraft, J., 2004. Phylogeny and diversification of the largest avian radiation. Proc. Natl. Acad. Sci. U.S.A. 101, 11040-11045.

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 Biol. Sci. 272, 849-858.

Biedler, J., Tu, Z. J., 2003. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. Mol. Biol. Evol. 20, 1811-1825.

Burch, J. B. E., Davis, D. L., Haas, N. B., 1993. Chicken repeat-1 elements contain a *pol*-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. Proc. Natl. Acad. Sci. U.S.A. 90, 8199-8203.

Chen, Z. Q., Ritzel, R. G., Lin, C. C., Hodgetts, R. B., 1991. Sequence conservation in avian CR-1 - an interspersed repetitive DNA family evolving under functional constraints. Proc. Natl. Acad. Sci. U.S.A. 88, 5814-5818.

Chesser, R. T., 2004. Molecular systematics of New World suboscine birds. Mol. Phylogenet. Evol. 32, 11-24.

Chikuni, K., Minaka, N., Ikenaga, H., 1996. Molecular phylogeny of some Passeriformes, based on cytochrome *b* sequences. J. Yamashina Inst. Ornithol. 28, 1-8.

Christy, B. A., Lau, L. F., Nathans, D., 1988. A gene activated in mouse 3T3-cells by serum growth-factors encodes a protein with zinc finger sequences. Proc. Natl. Acad. Sci. U.S.A. 85, 7857-7861.

Chubb, A. L., 2002. Molecular evolution and phylogenetic utility of the ZENK gene in birds. Dissertation. Department Integrative Biology, University of California, Berkeley.

Chubb, A. L., 2004a. New nuclear evidence for the oldest divergence among neognath birds: the phylogenetic utility of ZENK (i). Mol. Phylogenet. Evol. 30, 140-151.

Chubb, A. L., 2004b. Nuclear corroboration of DNA-DNA hybridization in deep phylogenies of hummingbirds, swifts, and passerines: the phylogenetic utility of ZENK (ii). Mol. Phylogenet. Evol. 30, 128-139.

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.

Clayton, D. F., 1997. Role of gene regulation in song circuit development and song learning. J. Neurobiol. 33, 549-571.

Cook, J. M., Tristem, M., 1997. 'SINEs of the times' - transposable elements as clade markers for their hosts. Trends Ecol. Evol. 12, 295-297.

Cracraft, J., 1992. Review: phylogeny and classification of birds. By Charles G. Sibley and Jon E. Ahlquist. Mol. Biol. Evol. 9, 182-186.

Cracraft, J., Feinstein, J., 2000. What is not a bird of paradise? Molecular and morphological evidence places *Macgregoria* in the Meliphagidae and the Cnemophilinae near the base of the corvoid tree. Proc. R. Soc. Lond. B Biol. Sci. 267, 233-241.

Cracraft, J., 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. Proc. R. Soc. Lond. B Biol. Sci. 268, 459-469.

Delacour, J., Amadon, D., 1951. The systematic position of *Picathartes*. Ibis 93, 60-62.

Drew, A. C., Brindley, P. J., 1997. A retrotransposon of the non-long terminal repeat class from the human blood fluke *Schistosoma mansoni*. Similarities to the chicken-repeat-1-like elements of vertebrates. Mol. Biol. Evol. 14, 602-610.

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.

Edwards, S. V., Arctander, P., Wilson, A. C., 1991. Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. Proc. R. Soc. Lond. B Biol. Sci. 243, 99-107.

Eickbush, T. H., 1992. Transposing without ends - the non-LTR retrotransposable elements. New Biologist 4, 430-440.

Ericson, P. G. P., Johansson, U. S., Parsons, T. J., 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, P. G. P., Christidis, L., Cooper, A., Irestedt, M., Jackson, J., Johansson, U. S., Norman, J. A., 2002a. A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. Proc. R. Soc. Lond. B Biol. Sci. 269, 235-241.

Ericson, P. G. P., Christidis, L., Irestedt, M., Norman, J. A., 2002b. Systematic affinities of the lyrebirds (Passeriformes: *Menura*), with a novel classification of the major groups of passerine birds. Mol. Phylogenet. Evol. 25, 53-62.

Ericson, P. G. P., Irestedt, M., Johansson, U. S., 2003. Evolution, biogeography, and patterns of diversification in passerine birds. J. Avian Biol. 34, 3-15.

Ericson, P. G. P., Johansson, U. S., 2003. Phylogeny of Passerida (Aves: Passeriformes) based on nuclear and mitochondrial sequence data. Mol. Phylogenet. Evol. 29, 126-138.

Fantaccione, S., Russo, C., Palomba, P., Rienzo, M., Pontecorvo, G., 2004. A new pair of CR1-like LINE and tRNA-derived SINE elements in *Podarcis sicula* genome. Gene 339, 189-198.

Farris, J. S., Kluge, A. G., Eckardt, M. J., 1970. A numerical approach to phylogenetic systematics. Syst. Zool. 19, 172-189.

Feduccia, A., 1975. Morphology of the bony stapes (columella) in the Passeriformes and related groups: evolutionary implications. Miscellaneous publication - University of Kansas, Museum of Natural History, 63, 1-34.

Feduccia, A., 1995. Explosive evolution in Tertiary birds and mammals. Science 267, 637-638.

Feduccia, A., 1996. The origin and evolution of birds, 2nd edition. Yale University Press, New Haven and London.

Felsenstein, J., 1981. Evolutionary trees from DNA-sequences - a maximum-likelihood approach. J. Mol. Evol. 17, 368-376.

Frith, C. B., Beehler, B. M., 1998. The birds of paradise. Oxford University Press, New York, NY.

Fuchs, J., Fjeldså, J., Bowie, R. C., 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.

Groth, J. G., 1998. Molecular phylogenetics of finches and sparrows: consequences of character state removal in cytochrome *b* sequences. Mol. Phylogenet. Evol. 10, 377-390.

Groth, J. G., Barrowclough, G. F., 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. 12, 115-123.

Haas, N. B., Grabowski, J. M., Sivitz, A. B., Burch, J. B. E., 1997. Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames. Gene 197, 305-309.

Haas, N. B., Grabowski, J. M., North, J., Moran, J. V., Kazazian, H. H., Burch, J. B. E., 2001. Subfamilies of CR1 non-LTR retrotransposons have different 5 ' UTR sequences but are otherwise conserved. Gene 265, 175-183.

Haffer, J., 1991. Familie Sylviidae - Zweigsänger (Grasmücken und Verwandte). In: Glutz von Blotzheim, U. N., Bauer, K. M. (Eds.) Handbuch der Vögel Mitteleuropas, AULA-Verlag Wiebelsheim.

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

Helmbychowski, K., Cracraft, J., 1993. Recovering phylogenetic signal from DNA-sequences - relationships within the corvine assemblage (Class Aves) as inferred from complete sequences of the mitochondrial-DNA cytochrome-*b* gene. Mol. Biol. Evol. 10, 1196-1214.

Huchon, D., Nikaido, M., Okada, N., 2002. Use of SINE retroposon in phylogeny. M S-Medecine Sciences 18, 1276-1281.

Huelsenbeck, J. P., Rannala, B., Masly, J. P., 2000. Accommodating phylogenetic uncertainty in evolutionary studies. Science 288, 2349-2350.

ICGSC - International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695-716.

IHGSC - International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Irestedt, M., Johansson, U. S., Parsons, T. J., Ericson, P. G. P., 2001. Phylogeny of major lineages of suboscines (Passeriformes) analysed by nuclear DNA sequence data. J. Avian Biol. 32, 15-25.

Irestedt, M., Fjeldså, J., Johansson, U. S., Ericson, P. G. P., 2002. Systematic relationships and biogeography of the tracheophone suboscines (Aves: Passeriformes). Mol. Phylogenet. Evol. 23, 499-512.

Jønsson, K. A., Fjeldså, J., 2006. A phylogenetic supertree of oscine passerine birds (Aves: Passeri). Zool. Scr. 35, 149-186.

Kajikawa, M., Ohshima, K., Okada, N., 1997. Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. Mol. Biol. Evol. 14, 1206-1217.

Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M., Okada, N., 1991. Shaping and reshaping of salmonid genomes by amplification of transfer RNA-derived retroposons during evolution. Proc. Natl. Acad. Sci. U.S.A. 88, 2326-2330.

Kido, Y., Saitoh, M., Murata, S., Okada, N., 1995. Evolution of the active sequences of the HpaI short interspersed elements. J. Mol. Evol. 41, 986-995.

Lanfear, J., Jowett, T., Holland, P. W. H., 1991. Cloning of fish zinc-finger genes related to Krox-20 And Krox-24. Biochem. Biophys. Res. Commun. 179, 1220-1224.

Lanyon, S. M., 1988. The stochastic mode of molecular evolution - what consequences for systematic investigations. Auk 105, 565-573.

Lanyon, S. M., 1992. Review: phylogeny and classification of birds. A study in molecular evolution by Charles G. Sibley, Jon E. Ahlquist. Condor 94, 304-307.

Larget, B., Simon, D. L., 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16, 750-759.

Long, K. D., Salbaum, J. M., 1998. Evolutionary conservation of the immediate-early gene ZENK. Mol. Biol. Evol. 15, 284-292.

Lovette, I. J., Bermingham, E., 2000. C-*mos* variation in songbirds: molecular evolution, phylogenetic implications, and comparisons with mitochondrial differentiation. Mol. Biol. Evol. 17, 1569-1577.

Lowe, P. R., 1938. Some anatomical and other notes on the systematic position of the genus *Picathartes*, together with some remarks on the families Sturnidae and Eulabetidae. Ibis, 254-269.

Lum, J. K., Nikaido, M., Shimamura, M., Shimodaira, H., Shedlock, A. M., Okada, N., Hasegawa, M., 2000. Consistency of SINE insertion topology and flanking sequence tree: quantifying relationships among cetartiodactyls. Mol. Biol. Evol. 17, 1417-1424.

Lutz, S. M., Vincent, B. J., Kazazian, H. H., Balzer, M. A., Moran, J. V., 2003. Allelic heterogeneity in LINE-1 retrotransposition activity. Am. J. Hum. Genet. 73, 1431-1437.

Malik, H. S., Burke, W. D., Eickbush, T. H., 1999. The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. 16, 793-805.

Mathews, L. M., Chi, S. Y., Greenberg, N., Ovchinnikov, I., Swergold, G. D., 2003. Large differences between LINE-1 amplification rates in the human and chimpanzee lineages. Am. J. Hum. Genet. 72, 739-748.

Mau, B., Newton, M. A., 1997. Phylogenetic inference for binary data on dendograms using Markov chain Monte Carlo. J. Comput. Graph. Stat. 6, 122-131.

Mau, B., Newton, M. A., Larget, B., 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Biometrics 55, 1-12.

McLachlan, G. R., Liversidge, R., 1978. Roberts birds of South Africa, 3rd ed. Trustees of the John Voelcker Bird Book Fund, Cape Town.

MGSC - Mouse Genome Sequencing Consortium, 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520-562.

Milbrandt, J., 1987. A nerve growth-factor induced gene encodes a possible transcriptional regulatory factor. Science 238, 797-799.

Mindell, D. P., 1992. DNA-DNA hybridization and avian phylogeny. Syst. Biol. 41, 126-134.

Moore, W. S., 1995. Inferring phylogenies from MtDNA variation - mitochondrial-gene trees versus nuclear-gene trees. Evolution 49, 718-726.

Nieselt-Struwe, K., von Haeseler, A., 2001. Quartet-mapping, a generalization of the likelihood-mapping procedure. Mol. Biol. Evol. 18, 1204-1219.

Nikaido, M., Rooney, A. P., Okada, N., 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. U.S.A. 96, 10261-10266.

Nikaido, M., Matsuno, F., Hamilton, H., Brownell, R. L., Cao, Y., Ding, W., Zuoyan, Z., Shedlock, A. M., Fordyce, R. E., Hasegawa, M., Okada, N., 2001. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. Proc. Natl. Acad. Sci. U.S.A. 98, 7384-7389.

Nunn, G. B., Cracraft, J., 1996. Phylogenetic relationships among the major lineages of the birds-of-paradise (Paradisaeidae) using mitochondrial DNA gene sequences. Mol. Phylogenet. Evol. 5, 445-459.

Okada, N., 1991. SINEs. Cur. Opin. Genet. Dev. 1, 498-504.

Poulter, R., Butler, M., Ormandy, J., 1999. A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons. Gene 227, 169-179.

Raikow, R. J., 1982. Monophyly of the Passeriformes - test of a phylogenetic hypothesis. Auk 99, 431-445.

Rannala, B., Yang, Z. H., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304-311.

Ribeiro, S., Mello, C. V., 2000. Gene expression and synaptic plasticity in the auditory forebrain of songbirds. Learn. Mem. 7, 235-243.

Rokas, A., Holland, P. W. H., 2000. Rave genomic changes as a tool for phylogenetics. Trends Ecol. Evol. 15, 454-459.

Sanzo, M., Stevens, B., Tsai, M. J., Omalley, B. W., 1984. Isolation of a protein-fraction that binds preferentially to chicken middle repetitive DNA. Biochemistry 23, 6491-6498.

Sasaki, T., Takahashi, K., Nikaido, M., Miura, S., Yasukawa, Y., Okada, N., 2004. First application of the SINE (Short interspersed repetitive element) method to infer

phylogenetic relationships in reptiles: an example from the turtle superfamily Testudinoidea. Mol. Biol. Evol. 21, 705-715.

Sato, A., Tichy, H., O'hUigin, C., Grant, P. R., Grant, B. R., Klein, J., 2001. On the origin of Darwin's finches. Mol. Biol. Evol. 18, 299-311.

Schmitz, J., Ohme, M., Zischler, H., 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics 157, 777-784.

Sclater, W. L., 1930. Systema avium Aethiopicarum. Taylor and Francis, London.

Sefc, K. M., Payne, R. B., Sorenson, M. D., 2003. Phylogenetic relationships of African sunbird-like warblers: moho (*Hypergerus atriceps*), green hylia (*Hylia prasina*) and tit-hylia (*Pholidornis rushiae*). Ostrich 74, 8-17.

Sharpe, R. B., 1883. Catalogue of the birds in the British Museum, Vol. 7. British Museum (Natural History), London.

Shedlock, A. M., Milinkovitch, M. C., Okada, N., 2000. SINE evolution, missing data, and the origin of whales. Syst. Biol. 49, 808-817.

Shedlock, A. M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. BioEssays 22, 148-160.

Sheldon, F. H., Gill, F. B., 1996. A reconsideration of songbird phylogeny, with emphasis on the evolution of titmice and their sylvioid relatives. Syst. Biol. 45, 473-495.

Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I., Okada, N., 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388, 666-670.

Sibley, C. G., Ahlquist, J. E., 1990. Phylogeny and classification of birds: a study in molecular evolution. Yale University Press, New Haven, CT.

Sibley, C. G., Monroe, B. L., 1990. Distribution and taxonomy of birds of the world. Yale University Press, New Haven, CT.

Silva, R., Burch, J. B. E., 1989. Evidence that chicken CR-1 elements represent a novel family of retroposons. Mol. Cell. Biol. 9, 3563-3566.

St. John, J., Cotter, J. P., Quinn, T. W., 2005. A recent chicken repeat 1 retrotransposition confirms the Coscoroba-Cape Barren goose clade. Mol. Phylogenet. Evol. 37, 83-90.

Stork, O., Welzl, H., 1999. Memory formation and the regulation of gene expression. Cell. Mol. Life Sci. 55, 575-592.

Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. U.S.A. 94, 6815-6819.

Stumph, W. E., Kristo, P., Tsai, M. J., Omalley, B. W., 1981. A chicken middle-repetitive DNA-sequence which shares homology with mammalian ubiquitous repeats. Nucl. Acids Res. 9, 5383-5397.

Sukhatme, V. P., Cao, X. M., Chang, L. C., Tsaimorris, C. H., Stamenkovich, D., Ferreira, P. C. P., Cohen, D. R., Edwards, S. A., Shows, T. B., Curran, T., Lebeau, M. M., Adamson, E. D., 1988. A zinc finger-encoding gene coregulated with c-*fos* during growth and differentiation, and after cellular depolarization. Cell 53, 37-43.

Swainson, W., 1832. Appendix, No. I; characters of genera and sub-genera hitherto undefined. *In* Swainson, W & Richardson, J. Fauna Boreali-Americana. Part 2, Birds. John Murray, London.

Thompson, H. S., Fotso, R., 1995. Rockfowl, the genus *Picathartes*. Bull. Afr. Bird Club 2, 25-28.

Tischmeyer, W., Grimm, R., 1999. Activation of immediate early genes and memory formation. Cell. Mol. Life Sci. 55, 564-574.

Van der Meij, M. A. A., de Bakker, M. A. G., Bout, R. G., 2005. Phylogenetic relationships of finches and allies based on nuclear and mitochondrial DNA. Mol. Phylogenet. Evol. 34, 97-105.

Vandergon, T. L., Reitman, M., 1994. Evolution of chicken repeat-1 (CR-1) elements - evidence for ancient subfamilies and multiple progenitors. Mol. Biol. Evol. 11, 886-898.

Verneau, O., Catzeflis, F., Furano, A. V., 1998. Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. Proc. Natl. Acad. Sci. U.S.A. 95, 11284-11289.

Vincent, B. J., Myers, J. S., Ho, H. J., Kilroy, G. E., Walker, J. A., Watkins, W. S., Jorde, L. B., Batzer, M. A., 2003. Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. Mol. Biol. Evol. 20, 1338-1348.

Voelker, G., Spellman, G. M., 2004. Nuclear and mitochondrial DNA evidence of polyphyly in the avian superfamily Muscicapoidea. Mol. Phylogenet. Evol. 30, 386-394.

Watanabe, M., Nikaido, M., Tsuda, T. T., Inoko, H., Mindell, D. P., Murata, K., Okada, N., 2006. The rise and fall of the CR1 subfamily in the lineage leading to penguins. Gene 365, 57-66.

Webster, M. T., Axelsson, E., Ellegren, H., 2006. Strong regional biases in nucleotide substitution in the chicken genome. Mol. Biol. Evol. 23, 1203-1216.

7 Appendix

7.1 Article I:

SIMONE TREPLIN, ROMY SIEGERT, CHRISTOPH BLEIDORN, HAZELL SHOKELLU THOMPSON, ROGER FOTSO, AND RALPH TIEDEMANN.

Looking for the 'best' marker: songbird (Aves: Passeriformes) phylogeny based on sequence analyses of several unlinked nuclear loci.

Systematic Biology, submitted.

Abstract

While the monophyly of the largest avian order Passeriformes as well as its suborders suboscines (Tyranni) and oscines (Passeri) is well established, lower phylogenetic relationships of this fast radiated taxon have been a continuous matter of debate, especially within the suborder oscines. Many studies analysing phylogenetic relationships of Passeriformes using molecular markers have been published which led to a better resolved phylogeny. Conflicting hypotheses and still remaining uncertainties, especially within Passerida, have repeatedly stimulated further research with additional new markers. In the present study we used a combination of established molecular markers (RAG-1, RAG-2, c-*myc*) and the recently introduced ZENK. We accomplished phylogenetic analyses using MP, ML, and Bayesian inference, both separately for all genes and simultaneously. To assess the phylogenetic utility of the different genes in avian systematics we analysed the influence of each data partition on the phylogenetic tree yielded by the combined approach using partitioned Bremer support. Compared to the other single gene analyses, the ZENK trees exhibited by far the best resolution and showed the lowest amount of homoplasy. Our data indicate that this gene is – at least in passerines – suitable for inference of even old taxonomic splits, dating before the Cretaceous/Tertiary boundary.

Our combined analysis yields well-supported phylogenetic hypotheses for passerine phylogeny and apart from corroborating recently proposed hypotheses on phylogenetic relationships in Passeriformes we provide evidence for some new hypotheses. The main passerine clades suboscines and oscines are corroborated just as paraphyly of Corvida. Based on the present study, we suggest a revision of the taxa Corvidae and Corvinae, as vireos are closer related to crows, ravens, and allies. The subdivision of Passerida into three superfamilies, Sylvioidea, Passeroidea, and Muscicapoidea, the first as sister to the two latter groups is strongly supported. We found evidence for a split within Sylvioidea into two clades, one consisting of tits and the other comprising warblers, bulbuls, laughingthrushes, whitethroats, and allies. Whereas Passeridae appear paraphyletic, monophyly of weavers and estrild finches as a separate clade is strongly supported. The sister taxon relationships of dippers and Muscicapidae is corroborated and we suggest a closer relationship of waxwings and kinglets to wrens, tree-creepers, and nuthatches.

Keywords: Passeriformes, phylogeny, nuclear markers, ZENK

INTRODUCTION

With more than 5700 species, the order Passeriformes comprises more than half of all living birds (Sibley and Ahlquist, 1990). The monophyly of this phenotypically rather homogenous taxon as well as its basal split into two monophyletic clades, i.e. the suboscines (Tyranni) and the oscines (Passeri), is well established on morphological (Ames, 1971; Feduccia, 1975; Raikow, 1982; Raikow, 1987) as well as on molecular grounds (e.g. Edwards et al., 1991). However, due to the rapid radiation of most passerine lineages during the early Tertiary (Feduccia, 1995), the phylogenetic relationships especially within the oscines have been a continuous matter of debate. The first extensive study on avian phylogenetic relationships, based on DNA-DNA hybridization (Sibley and Ahlquist, 1990) divided the oscines into the two parvorders Corvida and Passerida, the latter of which consists of the three superfamilies Muscicapoidea, Sylvioidea and Passeroidea. While subsequent sequence analyses generally have corroborated the suboscines/oscines partition (e.g. Edwards et al., 1991), the taxon Corvida is apparently paraphyletic and should not longer be maintained (Barker et al., 2002; Ericson et al., 2002a, b). In addition, further conflicting phylogenetic hypotheses have been put forward for lower phylogenetic relationships within the oscines (e.g. Ericson et al., 2003; Ericson and Johansson, 2003; e.g. Barker et al., 2004; Beresford et al., 2005). Recent studies on the systematics of passerine birds typically differ in their taxonomic sampling (with up to 173 passerine taxa included in Beresford et al. (2005)), but they generally rely on one or a few nuclear genes as phylogenetic markers, especially RAG-1 (Irestedt et al., 2001; Barker et al., 2002; Irestedt et al., 2002; Ericson and Johansson, 2003; Barker et al., 2004; Beresford et al., 2005), RAG-2 (Barker et al., 2004), c-*myc* (e.g. Ericson et al., 2000; Irestedt et al., 2001; Irestedt et al., 2002; e.g. Ericson and Johansson, 2003), and myoglobin (Irestedt et al., 2002; Ericson and Johansson, 2003).

In two recent studies, Chubb (2004a, b) demonstrated the utility of ZENK as a marker for a higher level phylogeny in neognath birds as well as for small analyses at a lower phylogenetic level of the avian orders Apodiformes (hummingbirds and swifts) and Passeriformes. While this author provided evidence that ZENK could be used as a powerful molecular marker with an estimated resolution for deep divergences within orders ranging roughly from 60 to 10 Mya, only 18 passerine taxa were included, such that an in depth analysis of phylogenetic relationships in this speciouse taxon was not possible.

ZENK (*sensu* Mello et al., 1992), encoded by an immediate-early gene (IEG), is a well-studied transcription factor expressed in the song system of birds (reviewed in Clayton,

1997; Ribeiro and Mello, 2000). ZENK is an acronym derived from the first character in the names of already described mammalian IEG homologs, i.e., **Z**if268 (Christy et al., 1988), **E**gr-1 (Sukhatme et al., 1988), **N**gfi-a (from humans, Milbrandt, 1987) and **K**rox-24 (Lanfear et al., 1991), all of which share conserved sequence elements (Long and Salbaum, 1998). Expression of ZENK plays an important role in neuronal growth concerning learning and memory formation (Stork and Welzl, 1999; Tischmeyer and Grimm, 1999; reviewed in Ribeiro and Mello, 2000) and has been used as a marker of neuro-activity during song learning and production (reviewed in Clayton, 1997; Ball and Gentner, 1998). Although the conservation of this single-copy gene and parts of its 3' untranslated region (UTR) is known since 1998 (Long and Salbaum, 1998), its use as a molecular marker in avian phylogenetics has been very limited so far.

In this study, we present a combined data set comprising 80 taxa sequenced at four nuclear genes. While three of them have already been applied on larger analyses of passerine phylogenies (RAG-1, RAG-2, c-*myc*, rfs. see above), we put emphasis on the newly introduced molecular marker ZENK. One aim was to verify the utility of ZENK as a marker for phylogenetic analyses. Second, we intended to re-assess previously proposed phylogenies within passerine birds. The phylogenetic reconstruction of relationships among closely related taxa specifically benefits from comprehensive data sets and a combination of unlinked markers. We accomplished analyses both separately for all genes and simultaneously for the combined data set, tested for homogeneity of the different gene combinations, and analysed the influence of each data partition on the phylogenetic tree yielded by the combined approach. This comprehensive analysis not only yields well-supported phylogenetic hypotheses for passerine phylogeny, but also assesses the phylogenetic utility of the different genes in avian systematics.

MATERIALS AND METHODS

Taxon Sampling, DNA Isolation, and Sequencing

Our taxon sampling of the combined data set comprised 80 taxa in total, representing 29 of the 46 passerine families recognised by Sibley and Monroe (1990) with emphasis on Passerida (*sensu* Sibley and Ahlquist, 1990). We here provide 184 new sequences (71 ZENK, 21 RAG-1, 27 RAG-2, 65 c-*myc*; GenBank, accession nos. **XXXXXX** – **XXXXXX)** and complement our data set with additional sequences available in GenBank (Table 1).

Total genomic DNA was extracted from blood using the DNeasy Tissue kit (Qiagen) and from liver tissue using the GNome® DNA Isolation Kit (Qbiogene). Additionally to previous published primers of ZENK and the 3' UTR (Chubb, 2004a), RAG-1 (Irestedt et al., 2001), RAG-2 (Barker et al., 2004), and c-*myc* (Ericson et al., 2000), new primers were developed to facilitate amplification of these four genes in passerines (Table 2). PCR-amplifications were performed in a total volume of 37.5 µl, containing 1mM Tris-HCl, pH 9.0, 5 mM KCl, 0.15 mM $MgCl_2$, 0.05 mM of each dNTP, 0.13 µM of both forward and reverse primers and 0.75 U *Taq* polymerase (Qbiogene) in two types of thermocyclers (Biometra, Biorad) according to the following reaction profiles: 1 cycle at 96°C for 5 min, 40 cycles at 96°C for 1 min 30 s, 51°C to 61 °C (depending on the primers' melting temperatures) for 1 min 15 s, 72°C for 1 min 30 s, and a final extension at 72°C for 10 min. Cycle sequencing reactions were performed with the forward and reverse primers using the BigDye version 3.1 Terminator Cycle Sequencing Kit (Applied Biosystems) and analysed on an AB 3100 multicapillary automatic sequencer (Applied Biosystems).

Phylogenetic Analysis

Sequences were assembled and aligned with the BioEdit Sequence Alignment Editor (Hall, 1999). Due to amplification with internal passerine-specific primers (see above), new sequences of RAG-1 and RAG-2 were shorter (1,420 bp and 847 bp, respectively) in length than some of the previously published ones (up to 2,887 bp and 1,152 bp, respectively). Missing nucleotides were treated as unknown. A chi-square test of homogeneity of base frequencies across taxa was used for each gene to test for variation of the base frequencies between the OTUs. Data sets of the different genes were tested for heterogeneity using the partition homogeneity test (Farris et al., 1995), implemented in PAUP* (Swofford, 2001), to assess the appropriateness of combining the data sets. We conducted a test between each pair of gene partitions using 1,000 replicates for each test.

All five data sets (i.e., each separate gene and the combined data set) were analysed by using maximum parsimony (MP, Farris et al., 1970), maximum likelihood (ML, Felsenstein, 1981) and Bayesian inference (Rannala and Yang, 1996; Mau and Newton, 1997; Larget and Simon, 1999; Mau et al., 1999; Huelsenbeck et al., 2000) with the representatives of the Old World suboscines (*Pitta sordida*, *Psarisomus dalhousiae*, and *Calyptomena viridis*) chosen as outgroups. MP analyses were performed using the parsimony ratchet approach (Nixon, 1999) as implemented in PAUPRat (Sikes and Lewis, 2001) and PAUP* (Swofford, 2001). The

ratchet search spanned 500 iterations, each of which included a unique and randomised weighting scheme, one random addition event, and TBR branch swapping. The search was repeated 20 times, because it is preferable to independently repeat the ratchet search rather than increasing the number of iterations (Nixon, 1999; Sikes and Lewis, 2001). All 10,020 trees were combined and only the best trees (ZENK: 8,722, RAG-1: 8,100, RAG-2: 9,526, c-*myc*: 1,260, combined data set: 10,020) were used to compute a consensus tree.

Additionally, a maximum parsimony bootstrap analysis with 1,000 iterations was performed for the combined data set using PAUP* (Swofford, 2001). With respect to time-consuming calculations, the maximum number of trees was limited to 100 and only one tree held at each step during stepwise addition.

For all five data sets, we used the Akaike Information Criterion (AIC), which is supposed to be superior to the hierarchical likelihood ratio test (Posada and Buckley, 2004) for model selection, as implemented in the program Modeltest version 3.7 (Posada and Crandall, 1998) for the ML analyses of the combined data set and as in MrModeltest version 2.2 (Nylander, 2004) for Bayesian inference.

ML analyses were performed under the likelihood settings suggested for the given dataset by the result of the modeltest (see Table 3) using PHYML (Guindon and Gascuel, 2003). The resulting likelihood tree was used as starting tree for a ML analyses using PAUP* (Swofford, 2001) with TBR branch swapping.

Bayesian analyses were conducted with a parallel version of MrBayes version 3.1 (Huelsenbeck and Ronquist, 2001). All priors were set according to the chosen model. Four Markov chains, three heated and one cold, were started from a random tree and all four chains ran simultaneously for 5,000,000 and 1,000,000 generations (for the combined analysis and for each separate gene, respectively), with trees being sampled every 500 and 100 generations for a total of 10,001 trees. After the likelihood of the trees of each chain converged, the first trees were discarded as *burn in* (6,001 combined, 600 ZENK and RAG-2, 400 RAG-1, 350 c-*myc*). The majority-rule consensus tree containing the posterior probabilities of the phylogeny was determined afterwards.

We estimated partitioned Bremer Support (PBS, Baker and DeSalle, 1997; Baker et al., 1998; Baker et al., 2001) for the strict consensus tree yielded from the MP analysis of the combined data set to assess the contribution of each gene to any node of the combined tree. PBS values were calculated using TreeRot version 2 (Sorenson, 1999) and PAUP* (Swofford, 2001). Where different analyses criteria yielded incongruent results, significance tests using

both the 'Approximately Unbiased' (AU) and the non-scaled bootstrap probability (NP) test of a tree topology selection provided by CONSEL (Shimodaira and Hasegawa, 2001; Shimodaira, 2002) were performed under the ML criterion for the combined data set for several *a priori* hypotheses (see results) against the best tree.

RESULTS

The complete combined data set consisted of 6,179 bp in total and was assembled of 1,651 bp ZENK (1,149 bp of exon 2 and 502 bp of 3'UTR), 2,887 bp RAG-1, 1,152 bp RAG-2, and 489 bp c-*myc*. Sequences varied in length due to several indels. Summary statistics for each gene and the combined data set are shown in Table 3. The number of parsimonious informative sites ranged from 89 (18.2 % of the complete sequence) in c-*myc* to 779 (27.0 %) in RAG-1. The chi-square test of homogeneity of base frequencies across taxa showed no significant heterogeneity (p=1.0) in all data sets. Sequences of ZENK had the lowest transition:transversion ratio (2.75) and the lowest proportion of invariable sites (0.252) whereas c-*myc* had the highest values of both criteria (4.44 and 0.53, respectively). All four data sets were tested for saturation. Transitions and transversions were plotted against genetic distance, but no saturation effects could be detected (data not shown).

Figure 1 shows the Bayesian phylogenetic tree of the ZENK data set and indicates those nodes corroborated by MP analysis. Generally, both MP and Bayesian analyses yield mostly congruent phylogenetic hypotheses. The basal split into suboscines and oscines was strongly supported. At the basis of the suborder oscines, the single representative of the honeyeaters Meliphagidae, the blue-faced honeyeater *Entomyzon cyanotis* was identified as the sister taxon to all remaining oscines. Picathartidae (genera *Picathartes* and *Chaetops*) branched off between Corvoidea (Corvidae + Vireonidae) and Passerida as sister group of the latter taxon. The three superfamilies of Passerida recognised by Sibley and Ahlquist (1990) were resolved with high Bayesian and parsimony support with the exception to the group of the wrens, tree-creepers and nuthatches (Certhiidae (*Troglodytes* and *Certhia*) and Sittidae). According to our analysis, this group is related to the Bohemian waxwing *Bombycilla garrulus* and the goldcrest *Regulus regulus.* Together, they represented the sister taxon to Passeroidea in this phylogenetic tree. Within Passeroidea, the monophyly of the fairy-bluebirds and leafbirds (Irenidae, represented by the Asian fairy-bluebird *Irena puella* and the greater green leafbird *Chloropsis sonnerati*) was not resolved. The same holds true for the

leaf-warblers (Acrocephalinae) within Sylvioidea (represented by the Blyth's reed-warbler *Acrocephalus dumetorum*, the icterine warbler *Hippolais icterina*, the willow warbler *Phylloscopus trochilus*, and the common grashopper-warbler *Locustella naevia*). Interestingly, both these contradictions between the 'classical' phylogeny (i.e., monophyly of Irenidae and Acrocephalinae) and our Bayesian analysis of the ZENK data set did not yield high Bayesian support and were not supported in the maximum parsimony analysis.

The partition homogeneity test showed no significant heterogeneity between all data partitions in the combined data set of all genes (Table 4). Figures 2 to 4 show the phylogenetic trees of the combined analyses (MP bootstrap analysis, ML with Bayesian support added at the nodes and strict consensus of the MP ratchet with partitioned Bremer support included, respectively). All trees showed high congruence and differed only in a few details, such as the position of wrens, tree-creepers and nuthatches, waxwings (Bombycillidae), kinglets (Regulidae), and leaf-warblers. Regarding the major groups of Passeriformes (i.e., suboscines/oscines, 'Corvida'/Passerida, Picathartidae, superfamilies of Passerida) all methods of analysing the combined dataset identified phylogenetic relationships similar to those with both high Bayesian and MP support in the analysis of the ZENK gene only. All taxa of Sylvioidea (*sensu* Sibley and Ahlquist, 1990) included in this study except the single representative of the larks, the Eurasian skylark *Alauda arvensis*, are the sister taxon to all Muscicapoidea and Passeroidea. The latter are clearly defined as a monophyletic group, whereas the relationships of the Muscicapoidea (*sensu* Sibley and Ahlquist, 1990) to the wrens and tree-creepers/nuthatches clade and the waxwings and kinglets remain ambiguous.

Of all MP strict consensus trees, the combined data set yielded the highest number of resolved nodes (71), followed by the MP strict consensus tree of ZENK with resolved nodes (51), much more than the other three separate partition analyses (8 to 23 nodes resolved, Table 3). Additionally, the ZENK tree comprises far more nodes that are congruent with the strict consensus of the combined data set than the other three genes (40, compared to 6-18, Table 3).

Several nodes are specifically discussed below and the Partitioned Bremer Support (PBS) values for these nodes are shown in Figure 5 (the complete PBS values for all 71 nodes across all partitions are available as supplementary material, online).The PBS values indicated that ZENK and RAG-1 had contributed to most of the nodes of the maximum parsimony strict consensus tree of the combined data set (supplementary material). A strong bias in resolving

phylogenies within Passeroidea with ZENK having the highest PBS values among all nodes was observed (supplementary material).

We tested seven different *a priori* phylogenetic hypotheses by AU and NP significance tests, i.e., we tested the tree constrained by a given hypothesis against the best tree (Table 5). For five hypotheses, the resulting tree did not significantly differ in topology from the best tree. Two *a priori* hypotheses were rejected, i.e., (a) the monophyly of Corvidae and Picathartidae and (b) the monophyly of Corvini + Oriolini + Vireonidae (except *Gymnorhina*).

DISCUSSION

Utility of Different Genes in Resolving Passerine Phylogeny

The enormous radiation of passerine birds has created a scenario where - 'instead of distinct evolutionary lines that can be traced by conventional methods, passerine phylogenies look like an upended head of an artist's camel hair paintbrush with the myriad single strands inextricably mixed' (Feduccia, 1996). As a result, there is a 'long history of frustration in oscine phylogenetics and classifications' (Sheldon and Gill, 1996). In recent years, many studies analysing phylogenetic relationships within Passeriformes have been published, leading step-by-step to an at least partially better resolved oscine phylogeny. However, conflicting hypotheses, especially within Passerida, have repeatedly reinforced our awareness of the obvious need for additional new markers to clarify these uncertainties. In the present study we tested the phylogenetic utility of ZENK with a significantly larger data set than before (Chubb, 2004b) and in combination with three other nuclear genes.

The separate analyses of all four partitions, in both MP and Bayesian inference, showed the highest resolution in the phylogenetic tree of ZENK, whereas single-gene trees of RAG-1, RAG-2, and especially c-*myc* suffered from more or less unresolved nodes. C-*myc* has been reported to be highly conserved throughout the vertebrate genome (Ericson et al., 2000). Indeed, our study indicates that this gene – with only 18.2 % of parsimonious informative sites – seems to be too conserved for resolving the phylogeny of a speciouse group like the passerines. Former studies using this gene used indels rather than sequence information for phylogenetic inference (Ericson et al., 2000) or analysed data sets with by far smaller taxon samplings and in combination with further molecular markers (Ericson et al.,

2002a, b; Ericson and Johansson, 2003). Surprisingly, despite of its low number of phylogenetically informative sites in our data set, this partition contributed most to the resolution of closely related taxa, i.e., at and near the genus-level (see Figs. 4 & 5 and Supplementary Material). The phylogenetic utility of RAG-1 has first been shown for basal divergences in birds in general (Groth and Barrowclough, 1999). RAG-1 was considered a powerful molecular marker for phylogenetic analyses among avian families. It is frequently included as an additional marker in passerine phylogenies based on more than one locus and dealing with varying systematic levels (Irestedt et al., 2001; Barker et al., 2002; Ericson et al., 2002a, b; Irestedt et al., 2002; Johansson et al., 2002; Ericson and Johansson, 2003; Barker et al., 2004; Cibois and Cracraft, 2004; Beresford et al., 2005). Rapidly evolving taxa were harder to resolve with RAG-1 (Groth and Barrowclough, 1999; Irestedt et al., 2001) and most of the published phylogenetic trees lack robust support for several nodes. High resolution of the phylogenetic relationships within suboscines was gained from analyses including RAG-1 (Irestedt et al., 2001; Irestedt et al., 2002; Barker et al., 2004; Beresford et al., 2005). This corroborates the phylogenetic utility of RAG-1 for deeper divergences, as the suboscines are less speciouse and comprise less closely related evolutionary lineages than the oscines (Sibley and Ahlquist, 1990). In the present study, the RAG-1 data set had low power in resolving phylogenetic relationships of closely related taxa and the resolved nodes were not highly supported above the genus-level. Nevertheless, our analysis of PBS values showed that RAG-1 makes a strong contribution to the combined data set analyses (see Supplementary Material). RAG-2 has so far only been used as a phylogenetic marker in passerine phylogenetics in combination with RAG-1. As a consequence, the phylogenetic utility of RAG-2 alone has never been evaluated. Although RAG-2 contains the highest proportion of parsimonious informative characters, only 8 nodes (the lowest value of all genes) are resolved in the MP strict consensus tree (Table 3). This marker performed poorly in resolving the passerine superfamilies and MP and Bayesian analyses were particularly incongruent when based on RAG-2. The contribution of RAG-2 to combined analyses as indicated by PBS values is much lower than those of RAG-1 are (see Supplementary Material). Compared to the other single gene analyses, the ZENK trees exhibited by far the best resolution with only a few inconsistencies between the Bayesian and the MP tree beyond family level (Fig. 1). Test for homoplasy using CI, RI and RC indices showed the lowest amount of homoplasy in the ZENK data set compared to the other three genes (Table 3), which further adds to its superior ability to resolve passerine phylogenies. According to Chubb (2004b), the highest power of ZENK is in resolving lineages which diverged roughly from 60 to 10 Mya ago. Our data

indicate that this gene is – at least in passerines – suitable for inference of even older taxonomic splits, i.e., before the Cretaceous/Tertiary boundary: The split into the suboscine taxa of Furnarioidea and Tyrannoidea is estimated to 61-65 Mya and into the suborders suboscines and oscines to ~76 Mya (Barker et al., 2004), both resolved with strong support on our phylogenetic trees of the ZENK data set (Fig.1). The PBS values show a comparatively similar contribution as RAG-1, with a strong bias in resolving phylogenies among Passeroidea (see Supplementary Material).

Phylogenetic Implications

All analyses of the combined data set yielded highly resolved, strongly supported phylogenetic trees. Given our so far less complete taxon sampling of suboscine and non-Passerida oscines, our analysis is less conclusive for these taxa than for the phylogeny within Passerida. Nevertheless, our analysis further advances our knowledge on passerine phylogeny, both by corroborating existing taxa and forwarding new hypotheses.

Suboscines.–The partition of the New-World suboscines in three clades, Sibley and Ahlquist's (1990) Tyrannida, Furnariida typical antbirds (Thamnophilidae) is not corroborated by our analyses. Instead, the integration of the typical antbirds into Furnariida (Irestedt et al., 2001; Irestedt et al., 2002; Chesser, 2004), as well as monophyly of the ovenbirds and woodcreepers (Furnariidae) and their sister group relationship with the ground antbirds (Formicariidae) is supported, the latter by all different data sets (node 6 in Fig. 4).

'*Corvida*'.–The monophyly of 'Corvida', a taxon originally proposed by Sibley and Ahlquist (1990), has been doubted by several authors (Barker et al., 2002; Ericson et al., 2002a, b). Presently, 'Corvida' are generally considered paraphyletic. This is again confirmed by the results of our study, as the honeyeaters (originally included in the 'Corvida' branch) are identified as a sister taxon to all other oscines (node 70 in Fig. 4). Our analyses also suggest a further revision of relationships within the superfamily Corvoidea (*sensu* Sibley and Ahlquist, 1990): we corroborate the sister taxa relationship between birds of paradise (Paradisaeini, included taxon *Manucodia* sp.) and corvids (Corvini, *Cyanocitta* sp., *Pica pica* and *Corvus* spp.; node 67 in Fig. 4), as previously hypothesised (Helmbychowski and Cracraft, 1993; Nunn and Cracraft, 1996; Frith and Beehler, 1998; Cracraft and Feinstein, 2000). Our data sets does not add further evidence on the phylogenetic positions of the currawongs (Artamini, *Gymnorhina tibicen*), orioles (Oriolini, *Oriolus* sp.) and vireos (Vireonidae, *Vireo* sp.) within Corvoidea. The position of *Gymnorhina* relative to Corvidae was ambiguous in the MP

analyses, but the phylogenetic hypothesis excluding this taxon from the remaining Corvidae was rejected by the significance test (hypothesis 8 in Table 5). These analyses suggest that both Sibley and Ahlquist's (1990) definition of the family Corvidae and the subfamily Corvinae (i.e., excluding vireos) should not be maintained. On the contrary, vireos are apparently closely related to the corvids and might even be nested within Corvidae. This assemblage of Corvidae and Vireonidae is strongly supported by the PBS values of all four genes (Figs. 1-3, 4, node 68).

Picathartidae.–The genus *Picathartes* has long been regarded as an avian curiosity (Thompson and Fotso, 1995) and their phylogenetic position has been an ongoing puzzle. Many different classifications have been put forward based on morphological traits after their initial description as a crow (*Corvus gymnocephalus*, Temminck 1825): they were classified as babblers (Amadon, 1943; Delacour and Amadon, 1951), starlings (Lowe, 1938), corvids (Sclater, 1930) and thrushes (Amadon, 1943). Sibley and Ahlquist (1990) remained unsure about the phylogenetic position of *Picathartes* and *Chaetops*, the closest relative of the rockfowls, and granted them a separate parvorder with the status of *incertae sedis*, aside all other Passeri. The position of Picathartidae (*Picathartes gymnocephalus*, *P. oreas*, *Chaetops frenatus*, *C. aurantius*) at the boundary beside 'Corvida' and Passerida found in our analyses has been proposed by several authors before (Ericson and Johansson, 2003; Barker et al., 2004; Beresford et al., 2005), but never gained robust statistical support. Our analyses do support this position strongly (bootstrap value=90, Fig. 2, PBS support by all four genes, Fig. 4, node 63). Nevertheless, the discussion about this group and some others, which seem to constitute a deep split within Passerida, e.g. the genus *Hyliota* (Fuchs et al., 2006), did not come to rest yet. A recent analysis based on retrotransposon insertions (assumed to constitute a truly apomorphic molecular character state) provide strong evidence for Picathartidae being closer related to Corvidae (Treplin and Tiedemann, under review) than to Passerida. Unlike this finding, the significance test in the present study, where the included taxa of Corvidae and Picathartidae were constrained to a monophylum gained significantly less support than the best tree (hypothesis 7, Table 5). Though, analyses of our RAG-1 data set resulted in Picathartidae as a member of 'Corvida' as the earliest branch in oscines except the honeyeaters with significant Parsimony support. These findings and the additionally ambiguous position of Petroicidae (Jønsson and Fjeldså, 2006, Treplin and Tiedemann, under review) as a proposed sister-taxon of Picathartidae (Beresford et al., 2005), indicate persisting difficulties to clarify phylogenetic relationship at the boundary between 'Corvida' and Passerida. Hence, we conclude that alternative hypotheses about the phylogenetic

relationships of Picathartidae, especially with additional markers, should be seriously considered.

Passerida.–The partition of Passerida into three superfamilies Sylvioidea, Muscicapoidea and Passeroidea (originally defined by Sibley and Ahlquist (1990)), with the first being the sister taxon of the two latter groups, is strongly corroborated by our analyses, but with slight modifications (Figs. 1-3, 4, nodes 28, 41, 42, 59). Non-monophyly of Sylvioidea has been shown before and is also corroborated by our analyses. Our study shows that two of the four 'sylvioid' clades proposed by Alström et al. (2006) are closely related i.e. tits (Paridae) and Sylvioidea (*sensu* Alström et al., 2006). The exact phylogenetic position of Paridae has often not been resolved in previous, even recent studies (e.g. Alström et al., 2006). Our data strongly support the exclusion of Paridae from Sylvioidea (*sensu* Alström et al., 2006) as a separate clade at the basis of Sylvioidea (Figs. 1-3, 4, node 58, 59). In addition, further taxa (*Elminia*, *Culicicapa*, and *Stenostira*, all with different classifications (Monarchini, Eopsaltriidae, Acrocephalinae, respectively) by Sibley and Monroe (1990)) have been suggested as sister-taxa to Paridae (Barker et al., 2004; Beresford et al., 2005), excluded from Sylvioidea. To keep the denomination 'Sylvioidea' introduced by Alström et al. (2006) we suggest to assign the new superfamily name 'Paroidea' to the remaining clade including tits and their relatives (e.g. Aegithalidae). Although taxonomic categories (like superfamilies) are not based on absolute criteria, the denomination of the two strongly supported clades at node 59 in Fig. 4 as Sylvioidea and Paroidea might ease further discussion on their respective phylogenetic relationships. The two remaining 'sylvioid' clades proposed by Alström et al. (2006), i.e. the nuthatch/treecreeper/gnatcatcher/wren clade, (Certhioidea *sensu* Cracraft et al. (2004)) and the crest/kinglet clade are not closely related to Sylvioidea and will be discussed below.

Our data show some incongruence among the different analyses regarding the position of larks (*Alauda* sp.). In both likelihood analyses of the combined data set, *Alauda* is the sister taxon to the remaining Sylvioidea. In contrast, they form a monophylum with the swallows (Hirundininae) in both MP analyses, and RAG-2 and c-*myc* contributed to this clade, as indicated by the partitioned Bremer support (Fig. 4, node 46). Significance tests for these two possibilities showed no statistically significant differences (Table 5). However, there are no hints for a monophyletic clade consisting of larks and swallows from other studies. Instead, *Alauda* formed the deepest branch within Sylvioidea (Barker et al., 2004; Beresford et al., 2005), in some studies together with *Panurus* (Ericson and Johansson, 2003; Alström et al., 2006; Fuchs et al., 2006) which was not included in our study.

Another incongruence concerns leaf-warblers (Acrocephalidae, *sensu* Alström et al., 2006). The phylogenetic position of the common grashopper-warbler (*Locustella naevia*) could not be unambiguously clarified. It formed a monophylum with *Acrocephalus* and *Hippolais* in both MP analyses, but appeared as the basal sister taxon of Sylvioidea (excl. *Alauda*) in both ML analyses. Haffer (1991) suggested a close relationship among these three taxa. However, not only this relationship has been questioned (Helbig and Seibold, 1999), but also the monophyly of both *Acrocephalus* and *Hippolais* was challenged (Leisler et al., 1997; Helbig and Seibold, 1999). In our analysis, bootstrap support for the clade including these three taxa was very low (Fig. 2) and only two of four genes contributed to this clade (ZENK and RAG-2, Fig. 4, node 57). Again, no significant differences were detected in the significant test (Table 5), such that the phylogenetic position of *Locustella* remains uncertain.

Within Sylvioidea, we yielded high resolution and congruence for a clade consisting of bulbuls (Pycnonotidae, node 48 in Fig. 4) and Timaliidae (*sensu* Alström et al., 2006). This relationship is specifically supported by RAG-1 (Fig. 4, node 53). This clade has been found in other studies also using RAG-1 and includes Hypocolidae and Cisticolidae (Barker et al., 2002; Barker et al., 2004; Beresford et al., 2005). Generally, the phylogeny within Sylvioidea (*sensu* Alström et al., 2006) has been difficult to elucidate (Alström et al., 2006; Jønsson and Fjeldså, 2006) and this taxon has appeared to be the least resolved group in the recently published supertree of Passerida in a metaanalysis of 99 studies (Jønsson and Fjeldså, 2006). In this group, radiation and speciation might have been particularly rapid, as indicated by the very short branch lengths in our phylogenetic analyses (Fig. 2 and 3).

The assignment of taxa to our Passeroidea clade is in general agreement with earlier studies, apart from the inclusion of the fairy-bluebirds and the leafbirds (Irenidae), which were classified as ‘Corvida’ by Sibley and Ahlquist (1990), and the exclusion of *Alauda* (see above).

At lower levels our study clearly showed that Sibley and Ahlquist’s (1990) family Passeridae should not be maintained. Their family Fringillidae (except *Peucedramus*) appeared to be embedded in sparrows (Passerinae) and wagtails and pipits (Motacillinae). We found strong support for a split of Passeridae into two clades, one consisting of sparrows, wagtails, pipits, and Fringillidae and the other consisting of weavers (Ploceinae) and estrildine finches (Estrildinae) (all taxa *sensu* Sibley and Ahlquist, 1990). This split has been postulated earlier, even though this hypothesis was little supported by empirical evidence (Groth, 1998), and was recently corroborated based on a combined data set of cytochrome *b* and β-fibrinogen

sequences (Van der Meij et al., 2005). The monophyly of weavers and estrildine finches is significantly supported in our analyses by Bayesian and MP values and by the PBS values of all four partitions (Figs. 1-3, 4, node 25). Additionally, the so far controversial position of *Vidua* (Sibley and Ahlquist, 1990; Groth, 1998) as the basal branch of estrildine finches is strengthened (node 24 in Fig. 4). Whether *Prunella* is the sister taxon to these two larger groups or closer related to the sparrows (as indicated by the ZENK data set and the MP bootstrap analysis of the combined data set, Figs. 1 and 2, respectively) is not fully resolved by our analyses, due to some incongruence among the different analysis methods. The significance test for these two possible topologies found no significant difference (Table 5). *Prunella* is proposed to be allied with the olive warbler *Peucedramus* (Ericson and Johansson, 2003), which was not included in this study and was described to represent the deepest branch in the clade of Passeridae and Fringillidae (Ericson and Johansson, 2003; Barker et al., 2004; Beresford et al., 2005). We disagree with the supertree proposed by Jønsson and Fjeldså (2006), where *Peucedramus* and *Prunella* are more closely related to weavers and estrildine finches.

Our studies strongly corroborate recent findings of the phylogeny of Muscicapoidea. Especially the position of dippers (*Cinclus*) as sister taxon to Muscicapidae (*sensu* Sibley and Ahlquist, 1990) is strengthened for the first time by significant MP support (Figs. 2, 4, node 39), in addition to significant Bayesian support which had been found in previous studies (Barker et al., 2002; Barker et al., 2004; Cibois and Cracraft, 2004; Beresford et al., 2005). Some controversies have been discussed concerning the group of wrens, tree-creepers, and nuthatches (Certhiidae and Sittidae) which were placed to Sylvioidea by Sibley and Ahlquist (1990). Recent studies based on RAG-1 and RAG-2 sequences suggested a close relationship to Muscicapoidea, but failed to support a clear phylogenetic position (Barker et al., 2004; Cibois and Cracraft, 2004; Beresford et al., 2005). Jønsson and Fjeldså (2006) found them at the basis of the Muscicapoidea including *Bombycilla,* but considered the latter taxon a 'problem clade'. In fact, the phylogenetic position *Bombycilla* has not been consolidated in recent studies, where it is either not resolved at all (Ericson and Johansson, 2003; Fuchs et al., 2006), associated with *Parus* as the deepest branch within Sylvioidea (Barker et al., 2002), or basal to Muscicapoidea (Barker et al., 2004; Voelker and Spellman, 2004; Beresford et al., 2005). The study of Barker et al. (2004) showed an affinity of *Bombycilla* to the goldcrest (*Regulus*)*,* which was, however, not supported significantly. *Regulus* itself, classified as a member of Sylvioidea by Sibley and Ahlquist (1990), was recently called another 'lost lineage in the passerine tree (Jønsson and Fjeldså, 2006). In our analysis, the sister

relationship of *Bombycilla* and *Regulus* is corroborated, mainly by the ZENK data set but also by RAG-1 and c-*myc* data sets (Figs. 1, 4, node 29). The tested hypothesis that *Bombycilla* and *Regulus* form a monophylum, as suggested by the different MP analyses, showed no significant difference in likelihood. In addition, it appears possible that *Bombycilla* and *Regulus* form a clade together with the group of wrens, treecreepers, and nuthatches is proposed: Albeit not supported significantly by Bayesian and MP bootstrap values, this clade was resolved in the MP ratchet strict consensus tree and supported by ZENK and c-*myc* as indicated by PBS (Figs. 1-3, 4, node 32). Of all hypotheses tested, the constrained monophyly of these taxa showed the highest congruence in likelihood (Table 5). This new hypothesis of phylogenetic relationships among waxwings, kinglets, wrens, tree-creepers and nuthatches awaits further evaluation based on a detailed investigation with a denser taxon sampling within this group.

ACKNOWLEDGEMENTS

We thank the following institutions and people for providing needed samples: Transvaal Museum, South Africa (T. Cassidy), University of Kiel, Haustierkunde (H. Luttmann and J. Oesert), Zoological Garden Berlin (R. Reinhard), Tierpark Berlin (M. Kaiser), Vogelpark Walsrode (B. Marcordes), Eberhard Curio, Dieter Wallschläger, Tomek Janiszewski, and Malte Mews. We also thank Katja Moll for laboratory assistance. Financial support appeared from the University of Potsdam.

REFERENCES

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.

Amadon, D., 1943. The genera of starlings and their relationships. Am. Mus. Novit. 1247, 1-16.

Ames, P., 1971. The morphology of the syrinx in passerine birds. Bull. Peabody Mus Nat. Hist. 37, 194.

Baker, R. H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46, 654-673.

Baker, R. H., Yu, X. B., DeSalle, R., 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol. 9, 427-436.

Baker, R. H., Wilkinson, G. S., DeSalle, R., 2001. Phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae). Syst. Biol. 50, 87-105.

Ball, G. F., Gentner, T. Q., 1998. They're playing our song: gene expression and birdsong perception. Neuron 21, 271-274.

Barker, F. K., Barrowclough, G. F., Groth, J. G., 2002. A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. Proc. R. Soc. Lond. B Biol. Sci. 269, 295-308.

Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., Cracraft, J., 2004. Phylogeny and diversification of the largest avian radiation. Proc. Natl. Acad. Sci. U.S.A. 101, 11040-11045.

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 Biol. Sci. 272, 849-858.

Chesser, R. T., 2004. Molecular systematics of New World suboscine birds. Mol. Phylogenet. Evol. 32, 11-24.

Christy, B. A., Lau, L. F., Nathans, D., 1988. A gene activated in mouse 3T3-cells by serum growth-factors encodes a protein with zinc finger sequences. Proc. Natl. Acad. Sci. U. S. A. 85, 7857-7861.

Chubb, A. L., 2004a. New nuclear evidence for the oldest divergence among neognath birds: the phylogenetic utility of ZENK (i). Mol. Phylogenet. Evol. 30, 140-151.

Chubb, A. L., 2004b. Nuclear corroboration of DNA-DNA hybridization in deep phylogenies of hummingbirds, swifts, and passerines: the phylogenetic utility of ZENK (ii). Mol. Phylogenet. Evol. 30, 128-139.

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.

Clayton, D. F., 1997. Role of gene regulation in song circuit development and song learning. J. Neurobiol. 33, 549-571.

Cracraft, J., Feinstein, J., 2000. What is not a bird of paradise? Molecular and morphological evidence places *Macgregoria* in the Meliphagidae and the Cnemophilinae near the base of the corvoid tree. Proc. R. Soc. Lond. B Biol. Sci. 267, 233-241.

Cracraft, J., Barker, F. K., Braun, M., Harshman, J., Dyke, G. J., Feinstein, J., Stanley, S., Cibois, A., Schikler, P., Beresford, P., García-Moreno, J., Sorenson, M. D., Yuri, T., Mindell, D. P., 2004. Phylogenetic relationships among modern birds (Neornithes) - toward an avian tree of life. Oxford University Press.

Delacour, J., Amadon, D., 1951. The systematic position of *Picathartes*. Ibis 93, 60-62.

Edwards, S. V., Arctander, P., Wilson, A. C., 1991. Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. Proc. R. Soc. Lond. B Biol. Sci. 243, 99-107.

Ericson, P. G. P., Johansson, U. S., Parsons, T. J., 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, P. G. P., Christidis, L., Cooper, A., Irestedt, M., Jackson, J., Johansson, U. S., Norman, J. A., 2002a. A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. Proc. R. Soc. Lond. B Biol. Sci. 269, 235-241.

Ericson, P. G. P., Christidis, L., Irestedt, M., Norman, J. A., 2002b. Systematic affinities of the lyrebirds (Passeriformes: *Menura*), with a novel classification of the major groups of passerine birds. Mol. Phylogenet. Evol. 25, 53-62.

Ericson, P. G. P., Irestedt, M., Johansson, U. S., 2003. Evolution, biogeography, and patterns of diversification in passerine birds. J. Avian Biol. 34, 3-15.

Ericson, P. G. P., Johansson, U. S., 2003. Phylogeny of Passerida (Aves: Passeriformes) based on nuclear and mitochondrial sequence data. Mol. Phylogenet. Evol. 29, 126-138.

Farris, J. S., Kluge, A. G., Eckardt, M. J., 1970. A numerical approach to phylogenetic systematics. Syst. Zool. 19, 172-189.

Farris, J. S., Kallersjo, M., Kluge, A. G., Bult, C., 1995. Constructing a significance test for incongruence. Syst. Biol. 44, 570-572.

Feduccia, A., 1975. Morphology of the bony stapes (columella) in the Passeriformes and related groups: evolutionary implications. Miscellaneous publication - University of Kansas, Museum of Natural History, 63, 1-34.

Feduccia, A., 1995. Explosive evolution in Tertiary birds and mammals. Science 267, 637-638.

Feduccia, A., 1996. The origin and evolution of birds, 2nd edition. Yale University Press, New Haven and London.

Felsenstein, J., 1981. Evolutionary trees from DNA-sequences - a maximum-likelihood approach. J. Mol. Evol. 17, 368-376.

Frith, C. B., Beehler, B. M., 1998. The birds of paradise. Oxford University Press, New York, NY.

Fuchs, J., Fjeldså, J., Bowie, R. C., 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.

Groth, J. G., 1998. Molecular phylogenetics of finches and sparrows: consequences of character state removal in cytochrome *b* sequences. Mol. Phylogenet. Evol. 10, 377-390.

Groth, J. G., Barrowclough, G. F., 1999. Basal divergences 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.

Haffer, J., 1991. Familie Sylviidae - Zweigsänger (Grasmücken und Verwandte). In: Glutz von Blotzheim, U. N., Bauer, K. M. (Eds.) Handbuch der Vögel Mitteleuropas, AULA-Verlag Wiebelsheim.

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

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

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

Helmbychowski, K., Cracraft, J., 1993. Recovering phylogenetic signal from DNA-sequences - relationships within the corvine assemblage (Class Aves) as inferred from complete sequences of the mitochondrial-DNA cytochrome-*b* gene. Mol. Biol. Evol. 10, 1196-1214.

Huelsenbeck, J. P., Rannala, B., Masly, J. P., 2000. Accommodating phylogenetic uncertainty in evolutionary studies. Science 288, 2349-2350.

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

Irestedt, M., Johansson, U. S., Parsons, T. J., Ericson, P. G. P., 2001. Phylogeny of major lineages of suboscines (Passeriformes) analysed by nuclear DNA sequence data. J. Avian Biol. 32, 15-25.

Irestedt, M., Fjeldså, J., Johansson, U. S., Ericson, P. G. P., 2002. Systematic relationships and biogeography of the tracheophone suboscines (Aves: Passeriformes). Mol. Phylogenet. Evol. 23, 499-512.

James, H. F., Ericson, P. G. P., Slikas, B., Lei, F. M., Gill, F. B., Olson, S. L., 2003. *Pseudopodoces humilis*, a misclassified terrestrial tit (Paridae) of the Tibetan Plateau: evolutionary consequences of shifting adaptive zones. Ibis 145, 185-202.

Johansson, U. S., Irestedt, M., Parsons, T. J., Ericson, P. G. P., 2002. Basal phylogeny of the Tyrannoidea based on comparisons of cytochrome *b* and exons of nuclear c-*myc* and RAG-1 genes. Auk 119, 984-995.

Jønsson, K. A., Fjeldså, J., 2006. A phylogenetic supertree of oscine passerine birds (Aves: Passeri). Zool. Scr. 35, 149-186.

Lanfear, J., Jowett, T., Holland, P. W. H., 1991. Cloning of fish zinc-finger genes related to Krox-20 and Krox-24. Biochem. Biophys. Res. Commun. 179, 1220-1224.

Larget, B., Simon, D. L., 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16, 750-759.

Leisler, B., Heidrich, P., SchulzeHagen, K., Wink, M., 1997. Taxonomy and phylogeny of reed warblers (genus *Acrocephalus*) based on mtDNA sequences and morphology. Journal für Ornithologie 138, 469-496.

Long, K. D., Salbaum, J. M., 1998. Evolutionary conservation of the immediate-early gene ZENK. Mol. Biol. Evol. 15, 284-292.

Lowe, P. R., 1938. Some anatomical and other notes on the systematic position of the genus *Picathartes*, together with some remarks on the families Sturnidae and Eulabetidae. Ibis, 254-269.

Mau, B., Newton, M. A., 1997. Phylogenetic inference for binary data on dendograms using Markov chain Monte Carlo. J. Comput. Graph. Stat. 6, 122-131.

Mau, B., Newton, M. A., Larget, B., 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Biometrics 55, 1-12.

Mello, C. V., Vicario, D. S., Clayton, D. F., 1992. Song presentation induces gene-expression in the songbird forebrain. Proc. Natl. Acad. Sci. U.S.A. 89, 6818-6822.

Milbrandt, J., 1987. A nerve growth-factor induced gene encodes a possible transcriptional regulatory factor. Science 238, 797-799.

Nixon, K. C., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15, 407-414.

Nunn, G. B., Cracraft, J., 1996. Phylogenetic relationships among the major lineages of the birds-of-paradise (Paradisaeidae) using mitochondrial DNA gene sequences. Mol. Phylogenet. Evol. 5, 445-459.

Nylander, J. A. A., 2004. MrModeltest v2. Program distributed by the author.

Page, R. D. M., Holmes, E. C., 1998. Molecular evolution. A phylogenetic approach. Blackwell Science, Oxford.

Posada, D., Crandall, K. A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14 (9), 817-818.

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.

Raikow, R. J., 1982. Monophyly of the Passeriformes - test of a phylogenetic hypothesis. Auk 99, 431-445.

Raikow, R. J., 1987. Hindlimb myology and evolution of the old world suboscine passerine birds (Acanthisittidae, Pittidae, Philepittidae, Eurylaimidae). Ornithological monographs 41, 1-81.

Rannala, B., Yang, Z. H., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304-311.

Ribeiro, S., Mello, C. V., 2000. Gene expression and synaptic plasticity in the auditory forebrain of songbirds. Learn. Mem. 7, 235-243.

Sclater, W. L., 1930. Systema avium Aethiopicarum. Taylor and Francis, London.

Sheldon, F. H., Gill, F. B., 1996. A reconsideration of songbird phylogeny, with emphasis on the evolution of titmice and their sylvioid relatives. Syst. Biol. 45, 473-495.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246-1247.

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, CT.

Sibley, C. G., Monroe, B. L., 1990. Distribution and taxonomy of birds of the world. Yale University Press, New Haven, CT.

Sikes, D. S., Lewis, P. O., 2001. PAUPRat: PAUP* implementation of the parsimony ratchet. Distributed by the authors. Beta software, version 1. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, USA.

Sorenson, M. D., 1999. TreeRot, version 2. Boston University, Boston, MA.

Stork, O., Welzl, H., 1999. Memory formation and the regulation of gene expression. Cell. Mol. Life Sci. 55, 575-592.

Sukhatme, V. P., Cao, X. M., Chang, L. C., Tsaimorris, C. H., Stamenkovich, D., Ferreira, P. C. P., Cohen, D. R., Edwards, S. A., Shows, T. B., Curran, T., Lebeau, M. M., Adamson, E. D., 1988. A zinc finger-encoding gene coregulated with c-*fos* during growth and differentiation, and after cellular depolarization. Cell 53, 37-43.

Swofford, D. L., 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer Associates, Sunderland, MA.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. Math. Life Sci. 17, 57-86.

Thompson, H. S., Fotso, R., 1995. Rockfowl, the genus *Picathartes*. Bull. Afr. Bird Club 2, 25-28.

Tischmeyer, W., Grimm, R., 1999. Activation of immediate early genes and memory formation. Cell. Mol. Life Sci. 55, 564-574.

Van der Meij, M. A. A., de Bakker, M. A. G., Bout, R. G., 2005. Phylogenetic relationships of finches and allies based on nuclear and mitochondrial DNA. Mol. Phylogenet. Evol. 34, 97-105.

Voelker, G., Spellman, G. M., 2004. Nuclear and mitochondrial DNA evidence of polyphyly in the avian superfamily Muscicapoidea. Mol. Phylogenet. Evol. 30, 386-394.

Table 1 Taxa used in this study with accession numbers and references. Systematic classifications following Sibley and Monroe (1990). Reference for sequences published in GenBank: Ref. 1: Chubb 2004b, Ref. 2: Barker *et al*. 2004, Ref. 3: Barker *et al*. 2002, Ref. 4: Groth and Barrowclough 1999, Ref. 5: Irestedt et a. 2001, Ref. 6: Cibois and Cracraft 2004, Ref. 7: Barker *et al*. unpublished, Ref. 8: Ericson *et al*. 2002a, Ref. 9: Beresford *et al*. 2005, Ref. 10: Irestedt *et al*. 2002, Ref. 11: James *et al*. 2003.

	ZENK		RAG-1		RAG-2		c-*myc*	
Family - subfamily - tribe	species	Accession no.	species	Accession no.	species	Accession no.	species	Accession no.
Pittidae	*Pitta sordida*	this study	*Pitta sordida*	AY443319 Ref. 2	*Pitta sordida*	AY443206 Ref. 2	*Pitta sordida*	this study
Eurylaimidae	*Psarisomus dalhousiae*	AF492520 / AF492550 Ref. 1	*Psarisomus dalhousiae*	AY057025 Ref. 3	*Psarisomus dalhousiae*	AY443214 Ref. 2	*Calyptomena viridis*	AF295161 Ref. 5
Tyrannidae - Tyranninae	*Myiarchus cinerascens*	AF492517 / AF492547 Ref. 1	*Tyrannus tyrannus*	AF143739 Ref. 4	*Tyrannus tyrannus*	AY443243 Ref. 2	*Tyrannus savana*	AF295182 Ref. 5
Tyrannidae - Cotinginae	*Porphyrolaema porphyrolaema*	AF492519 / AF492549 Ref. 1	*Rupicola rupicola*	AY057029 Ref. 3	*Rupicola rupicola*	AY443224 Ref. 2	*Rupicola peruviana*	this study
	Procnias nudicollis	this study	*Procnias nudicollis*	this study	*Procnias nudicollis*	this study	*Procnias nudicollis*	this study
Tyrannidae - Piprinae	*Pipra coronata*	AF492518 / AF492548 Ref. 1	*Pipra coronata*	AY057020 Ref. 3	*Lepidothrix coronata*	AY443204 Ref. 2	*Pipra fasciicauda*	AF295175 Ref. 5
Thamnophilidae	*Thamnophilus caerulescens*	AF492521 / AF492551 Ref. 1	*Thamnophilus nigrocinereus*	AY057034 Ref. 3	*Thamnophilus nigrocinereus*	AY443235 Ref. 2	*Thamnophilus caerulescens*	AF295180 Ref. 5
Furnariidae - Furnariinae	*Furnarius leucopus*	this study	*Furnarius rufus*	AY056995 Ref. 3	*Furnarius rufus*	AY443149 Ref. 2	*Furnarius cristatus*	AF295165 Ref. 5
Furnariidae - Dendrocolaptinae	*Deconychura longicauda*	AF492515 / AF492545 Ref. 1	*Lepidocolaptes angustirostris*	AF295190 Ref. 5	*Campylorhamphus trochilirostris*	AY443112 Ref. 2	*Lepidocolaptes angustirostris*	AF295168 Ref. 5
Formicariidae	*Formicarius analis*	AF492516 / AF492546	*Formicarius colma*	AY056993	*Formicarius colma*	AY443147	*Formicarius nigricapillus*	AY065692

		Ref. 1		Ref. 3		Ref. 2		Ref. 10
Conopophagidae	*Conopophaga peruviana*	AF490188 / AF490243	*Conopophaga ardesiaca*	AY443271	*Conopophaga ardesiaca*	AY443125	*Conopophaga lineata*	AF295163
		Ref. 1		Ref. 2		Ref. 2		Ref. 5
Meliphagidae	*Entomyzon cyanotis*		*Meliphaga analoga*	AY057003	*Meliphaga analoga*	AY443170	*Ptiloprora plumbea*	AY037841
		this study		Ref. 3		Ref. 2		Ref. 8
Irenidae	*Irena puella*		*Irena cyanogaster*	AY056999	*Irena cyanogaster*	AY443158	*Irena puella*	
		this study		Ref. 3		Ref. 2		this study
	Chloropsis sonnerati		*Chloropsis cochinchinensis*	AY056984	*Chloropsis cochinchinensis*	AY443118	*Chloropsis sonnerati*	
		this study		Ref. 3		Ref. 2		this study
Vireonidae	*Vireo cassini*	AF492529 / AF492559	*Vireo philadelphia*	AY057041	*Vireo philadelphia*	AY443245	-	
		Ref. 1		Ref. 3		Ref. 2		
Corvidae - Corvinae - Corvini	*Corvus corax*		*Corvus corax*		*Corvus coronoides*	AY443133	*Corvus corax*	
		this study		this study		Ref. 2		this study
	Corvus corone		*Corvus corone*	AY056989	*Corvus corone*	AY443132	*Corvus corone*	
		this study		Ref. 3		Ref. 2		this study
	Cyanocitta stelleri		*Cyanocitta cristata*	AY443280	*Cyanocitta cristata*	AY443137	*Cyanocitta cristata*	
		this study		Ref. 2		Ref. 2		this study
	Pica pica		*Pica pica*		*Pica pica*		*Pica pica*	
		this study		this study		this study		this study
Corvidae - Corvinae - Paradisaeini	*Manucodia keraudrenii*		*Manucodia chalybata*	AY443296	*Manucodia chalybata*	AY443164	*Manucodia keraudrenii*	
		this study		Ref. 2		Ref. 2		this study
Corvidae - Corvinae - Artamini	*Gymnorhina tibicen*		*Gymnorhina tibicen*	AY443289	*Gymnorhina tibicen*	AY443153	*Gymnorhina tibicen*	
		this study		Ref. 2		Ref. 2		this study
Corvidae - Corvinae - Oriolini	*Oriolus chinensis*		*Oriolus xanthonotus*	AY443308	*Oriolus xanthonotus*	AY443185	*Oriolus chinensis*	
		this study		Ref. 2		Ref. 2		this study
Picathartidae	*Picathartes oreas*		*Picathartes*		-		*Picathartes oreas*	

			oreas					
		this study		this study				this study
	Picathartes gymnocephalus		*Picathartes gymnocephalus*	AY057019	*Picathartes gymnocephalus*	AY443203	*Picathartes gymnocephalus*	
		this study		Ref. 3		Ref. 2		this study
	Chaetops aurantius		*Chaetops frenatus*	AY443266	*Chaetops frenatus*	AY443116	*Chaetops aurantius*	
		this study		Ref. 2		Ref. 2		this study
Bombycillidae	*Bombycilla garrulus*		*Bombycilla garrulus*	AY056981	*Bombycilla garrulus*	AY443111	*Bombycilla garrulus*	
		this study		Ref. 3		Ref. 2		this study
Cinclidae	*Cinclus cinclus*		*Cinclus cinclus*	AY056985	*Cinclus cinclus*	AY443119	*Cinclus cinclus*	
		this study		Ref. 3		Ref. 2		this study
Muscicapidae - Turdinae	*Turdus philomelos*		*Turdus philomelos*	AY307214	*Turdus philomelos*		*Turdus philomelos*	
		this study		Ref. 6		this study		this study
	Turdus merula		*Turdus falcklandii*	AY057039	*Turdus falcklandii*	AY443242	-	
		this study		Ref. 3		Ref. 2		
	Zoothera naevia		*Zoothera dauma*	AY307215	*Zoothera naevia*		*Zoothera naevia*	
		this study		Ref. 6		this study		this study
	Catharus guttatus		*Catharus guttatus*	AY307184	*Catharus ustulatus*	AY443114	*Catharus guttatus*	
		this study		Ref. 6		Ref. 2		this study
Muscicapidae - Muscicapinae - Muscicapini	*Ficedula hypoleuca*		*Ficedula monileger*	AY307192	*Ficedula hypoleuca*		*Ficedula hypoleuca*	
		this study		Ref. 6		this study		this study
	Muscicapa striata		*Muscicapa ferruginea*	AY443305	*Muscicapa ferruginea*	AY443179	*Muscicapa striata*	
		this study		Ref. 2		Ref. 2		this study
Muscicapidae - Muscicapinae - Saxicolini	*Erithacus rubecula*		*Erithacus rubecula*	AY307191	*Erithacus rubecula*		*Erithacus rubecula*	
		this study		Ref. 6		this study		this study
	Saxicola rubetra		-		-		-	
		this study						

	Luscinia svecica	this study	*Luscinia cyane*	AY307196 Ref. 6	-		-	
	Phoenicurus ochruros	this study	*Phoenicurus frontalis*	AY307205 Ref. 6	*Phoenicurus ochruros*	this study	*Phoenicurus ochruros*	this study
Sturnidae - Sturnini	*Sturnus vulgaris*	this study	*Sturnus vulgaris*	AY057032 Ref. 3	*Sturnus vulgaris*	AY443232 Ref. 2	*Sturnus vulgaris*	this study
Sturnidae - Mimini	*Mimus polyglottos*	AF492525 / AF492555 Ref. 1	*Mimus patagonicus*	AY057005 Ref. 3	*Mimus patagonicus*	AY443173 Ref. 2	*Mimus saturninus*	AF377265 Ref. 11
Sittidae - Sittinae	*Sitta europaea*	this study	*Sitta pygmaea*	AY057030 Ref. 3	*Sitta carolinensis*	AY443227 Ref. 2	*Sitta europaea*	this study
Certhiidae - Certhiinae	*Certhia brachydactyla*	this study	*Certhia familiaris*	AY056983 Ref. 3	*Certhia familiaris*	AY443115 Ref. 2	*Certhia brachydactyla*	this study
Certhiidae - Troglodytinae	*Troglodytes troglodytes*	this study	*Troglodytes aedon*	AY057038 Ref. 3	*Troglodytes aedon*	AY443241 Ref. 2	*Troglodytes troglodytes*	this study
Paridae - Parinae	*Parus major*	this study	*Parus major*	AY443314 Ref. 2	*Parus major*	AY443197 Ref. 2	*Parus major*	this study
	Parus caeruleus	this study	*Parus inornatus*	AY057017 Ref. 3	*Parus inornatus*	AY443196 Ref. 2	*Parus caeruleus*	this study
	Parus cristatus	this study	*Parus cristatus*	this study	*Parus cristatus*	this study	*Parus cristatus*	this study
	Parus palustris	this study	*Parus palustris*	this study	*Parus palustris*	this study	*Parus palustris*	this study
Hirundinidae - Hirundininae	*Delichon urbica*	this study	*Hirundo pyrrhonota*	AY056997 Ref. 3	*Hirundo pyrrhonota*	AY443154 Ref. 2	*Delichon urbica*	this study
Regulidae	*Regulus regulus*	this study	*Regulus calendula*	AY057028 Ref. 3	*Regulus calendula*	AY443220 Ref. 2	*Regulus regulus*	this study
Pycnonotidae	*Hypsipetes philippinus*	this study	*Hypsipetes philippinus*	this study	*Hypsipetes philippinus*	this study	*Hypsipetes philippinus*	this study

	Pycnonotus leucogenys	this study	*Pycnonotus leucogenys*	this study	*Pycnonotus leucogenys*	this study	*Pycnonotus leucogenys*	this study
	Pycnonotus xanthopygos	this study	*Pycnonotus barbatus*	AY057027 Ref. 3	*Pycnonotus barbatus*	AY443219 Ref. 2	*Pycnonotus xanthopygos*	this study
Zosteropidae	*Zosterops senegalensis*	this study	*Zosterops senegalensis*	AY057042 Ref. 3	*Zosterops senegalensis*	AY443247 Ref. 2	*Zosterops senegalensis*	this study
Sylviidae - Acrocephalinae	*Acrocephalus dumetorum*	this study	*Acrocephalus newtoni*	AY319972 Ref. 7	*Acrocephalus newtoni*	AY799825 Ref. 9	*Acrocephalus dumetorum*	this study
	Hippolais icterina	this study	*Hippolais icterina*	this study	*Hippolais icterina*	this study	*Hippolais icterina*	this study
	Phylloscopus trochilus	this study	*Phylloscopus collybita*	AY319997 Ref. 7	*Phylloscopus collybita*	AY799844 Ref. 9	-	
	Locustella naevia	this study	*Locustella naevia*	this study	*Locustella naevia*	this study	*Locustella naevia*	this study
Sylviidae - Garrulacinae	*Garrulax sannio*	this study	*Garrulax milleti*	AY056996 Ref. 3	*Garrulax milleti*	AY443151 Ref. 2	-	
Sylviidae - Sylviinae - Timaliini	*Chamaea fasciata*	AF492523 / AF492553 Ref. 1	-		-		-	
Sylviidae - Sylviinae - Sylviini	*Sylvia atricapilla*	this study	*Sylvia atricapilla*	this study	*Sylvia atricapilla*	this study	*Sylvia atricapilla*	this study
	Sylvia communis	this study	*Sylvia communis*	this study	*Sylvia communis*	this study	*Sylvia communis*	this study
	Sylvia nisoria	this study	*Sylvia nana*	AY057033 Ref. 3	*Sylvia nana*	AY443233 Ref. 2	*Sylvia nisoria*	this study
Alaudidae	*Alauda arvensis*	this study	*Alauda arvensis*	AY056978 Ref. 3	*Alauda arvensis*	AY443106 Ref. 2	*Alauda arvensis*	this study
Nectariniidae -	*Nectarinia cuprea*		*Nectarinia*	AY057009	*Nectarinia olivacea*	AY443180	*Nectarinia cuprea*	

Nectariniinae				*olivacea*								
		this study			Ref. 3				Ref. 2			this study
Passeridae - Passerinae	*Passer domesticus*			*Passer domesticus*			*Passer domesticus*			*Passer domesticus*		
		this study			this study			this study				this study
	Passer montanus			*Passer montanus*	AF143738		*Passer montanus*	AY443198		*Passer montanus*		
		this study			Ref. 4			Ref. 2				this study
Passeridae - Motacillinae	*Motacilla alba*			*Motacilla cinerea*	AY057007		*Motacilla cinerea*	AY443178		*Motacilla alba*		
		this study			Ref. 3			Ref. 2				this study
Passeridae - Prunellinae	*Prunella modularis*			*Prunella collaris*	AY057024		*Prunella collaris*	AY443213		*Prunella modularis*		
		this study			Ref. 3			Ref. 2				this study
Passeridae - Ploceinae	*Dinemellia dinemelli*			*Dinemellia dinemelli*			*Dinemellia dinemelli*			*Dinemellia dinemelli*		
		this study			this study			this study				this study
	Ploceus vitellinus			*Ploceus cucullatus*	AY057022		*Ploceus cucullatus*	AY443207		*Ploceus vitellinus*		
		this study			Ref. 3			Ref. 2				this study
Passeridae - Estrildinae - Estrilidini	*Emblema pictum*			*Emblema pictum*			*Emblema pictum*			*Emblema pictum*		
		this study			this study			this study				this study
	Lonchura cantans			*Lonchura cantans*			*Lonchura cantans*			*Lonchura cantans*		
		this study			this study			this study				this study
	Mandingoa nitidula			-			*Mandingoa nitidula*			*Mandingoa nitidula*		
		this study						this study				this study
	Uraeginthus ianthinogaster			*Uraeginthus ianthinogaster*			*Uraeginthus ianthinogaster*			*Uraeginthus ianthinogaster*		
		this study			this study			this study				this study
	Estrilda nonnula			*Estrilda astrild*	AY319983		*Estrilda nonnula*			*Estrilda nonnula*		
		this study			Ref. 7			this study				this study
Passeridae - Estrildinae - Viduini	*Vidua macroura*			*Vidua macroura*			*Vidua macroura*			*Vidua macroura*		
		this study			this study			this study				this study

Fringillidae - Fringillinae - Fringillini	*Fringilla coelebs*		*Fringilla montifringilla*	AY056994	*Fringilla montifringilla*	AY443148	*Fringilla coelebs*	
		this study		Ref. 3		Ref. 2		this study
Fringillidae - Fringillinae - Carduelini	*Coccothraustes coccothraustes*		*Coccothraustes coccothraustes*	AY037855	*Coccothraustes coccothraustes*		*Coccothraustes coccothraustes*	
		this study		Ref. 8		this study		this study
	Carduelis carduelis		*Carduelis carduelis*		*Carduelis carduelis*		*Carduelis carduelis*	
		this study		this study		this study		this study
	Carduelis spinus		*Carduelis spinus*		*Carduelis spinus*		*Carduelis spinus*	
		this study		this study		this study		this study
Fringillidae - Emberizinae - Emberizini	*Passerella iliaca*		*Passerella iliaca*		*Passerella iliaca*		*Passerella iliaca*	
		this study		this study		this study		this study
	Emberiza citrinella		*Emberiza schoeniclus*	AY056992	*Emberiza schoeniclus*	AY443143	*Emberiza citrinella*	
		this study		Ref. 3		Ref. 2		this study
Fringillidae - Emberizinae - Thraupini	*Thraupis episcopus*		*Thraupis cyanocephala*	AY057035	*Thraupis cyanocephala*	AY443236	*Thraupis episcopus*	
		this study		Ref. 3		Ref. 2		this study

Table 2 Newly developed primers used in this study, primers 1-7 RAG-1, 8-12 RAG-2, and 13 ZENK.

R1L2	5' GTC CCC AAA CTG TGA TGT GTG C 3'
R1H3	5' GCA GTC TCG ATA AAA GGT TTG GC 3'
R1H4	5' GCA TTC ATG AAC TTC TGG AGG TA 3'
R1L3	5' GCC AGT AGA CAC AAT TGC AAA GAG 3'
R1L4	5' GTT TGT ACC CTG TGT GAT GCC AC 3'
R50int	5' GTC TGG CCA TCC GAA TCA ACA CGT TT 3'
R51int	5' CCT GAC AGT CCA TCT ATA ATT CCC AC 3'
R2K1int	5' GAC TTT CCT TCC ATG TTT CAA TTG C 3'
R2-O	5' GTT GAA AGT GTG AGC CCA GAG TGG AC 3'
R2-R	5' GAT GTA AAA GTA GTT TGC ATC TGG GCT 3'
R2R4int	5' GAG CCC CCA ACA AGG ACA AAT TC 3'
R2-V	5' GTG ACA TTC CAA TGC ATT GAG AAA GA 3'
Z7aR	5' GAA TGG CTT CTC TCC TGT GTG 3'

Table 3 Summary of sequence and MP trees for the separate genes and the combined data set.

	ZENK	RAG-1	RAG-2	c-*myc*	comb. data set
Size (bp)	1651	2887	1152	489	6179
PI[a]	422	779	363	89	1648
%G	18.2	24.2	23.4	24.9	22.4
%A	25.3	31.3	29.4	33.1	29.3
%T	26.5	24.1	26.0	17.5	24.5
%C	30.0	20.5	21.2	24.5	23.7
Ts/tv ratio	2.751	3.213	3.270	4.436	3.195
Model[b]	GTR	GTR	TVM	HKY	GTR
PINVAR[b]	0.252	0.374	0.296	0.531	0.344
Gamma[b]	0.835	1.149	0.916	0.577	0.943
CI[c]	0.546	0.513	0.495	0.436	0.506
CI of the strict consensus	0.525	0.360	0.304	0.237	0.503
RI[c]	0.642	0.581	0.607	0.647	0.596
RI of the strict consensus	0.610	0.214	0.115	0.122	0.591
RC[c]	0.350	0.298	0.301	0.282	0.302
RC of the strict consensus	0.320	0.077	0.035	0.029	0.298
Resolved nodes[d]	51	23	8	11	71
Congruent nodes[d]	40	18	6	8	-

[a] Parsimonious informative sites

[b] Models of molecular evolution represent the general time-reversible (GTR) model (Tavaré *et al*. 1986), transversion model (TVM) model (Posada and Crandall 1998), and the Hasegawa-Kishuno-Yano (HKY) model (Hasegawa *et al*. 1985) all both with assumptions of proportions of invariable sites (PINVAR) and gamma shape correction parameters (Page and Holmes 1998, Swofford, 2001).

[c] Measures of homoplasy (CI, RI, and RC values) are given for *n* equally parsimonious trees, followed by equivalent values for strict consensus.

[d] Resolved nodes give the number of completely resolved nodes, and congruent nodes shows the total number of resolved nodes, which are also present in the tree of the combined data set.

Table 4 Values of the homogeneity test for all combinations of the four nuclear genes.

	ZENK	RAG1	RAG2
RAG-1	0.240		
RAG-2	0.740	0.231	
c-*myc*	0.644	0.073	0.260

Table 5 Results of the significance tests. Obs: observed log-likelihood difference, au: p-value of the approximately unbiased test calculated from the multiscale bootstrap, np: bootstrap probability calculated from the multiscale bootstrap. All taxa refer to Sibley and Monroe (1990).

rank	tested hypotheses	supported by	obs	au	np
1	best tree		-	0.686	0.177
2	*Bombycilla*, *Regulus*, Sittidae and Certhiidae are monophyletic	MP ratchet analysis comb. data set, ZENK, c-*myc*	0.9	0.536	0.313
3	*Bombycilla* and *Regulus* form a monophylum	both MP comb. data set, ZENK, RAG-1, c-*myc*	0.9	0.536	0.313
4	*Prunella* + Passerinae + *Motacilla* + Fringillidae are monophyletic	MP bootstrap analysis comb. data set, ZENK	1	0.523	0.276
5	*Acrocephalus* + *Hippolais* + *Locustella* are monophyletic	both MP comb. data set	3.3	0.384	0.118
6	*Alauda* and Hirundininae are monophyletic	both MP comb. analysis, RAG-2, c-*myc*	9.5	0.229	0.117
7	Picathartidae form a monophylum with Corvidae	Treplin and Tiedemann (under review)	18.6	0.029	0.007
8	Corvini + Oriolini + Vireonidae except *Gymnorhina* are monophyletic	MP ratchet, comb. data set, RAG-2, c-*myc*	75.7	$7x10^{-5}$	$9x10^{-6}$

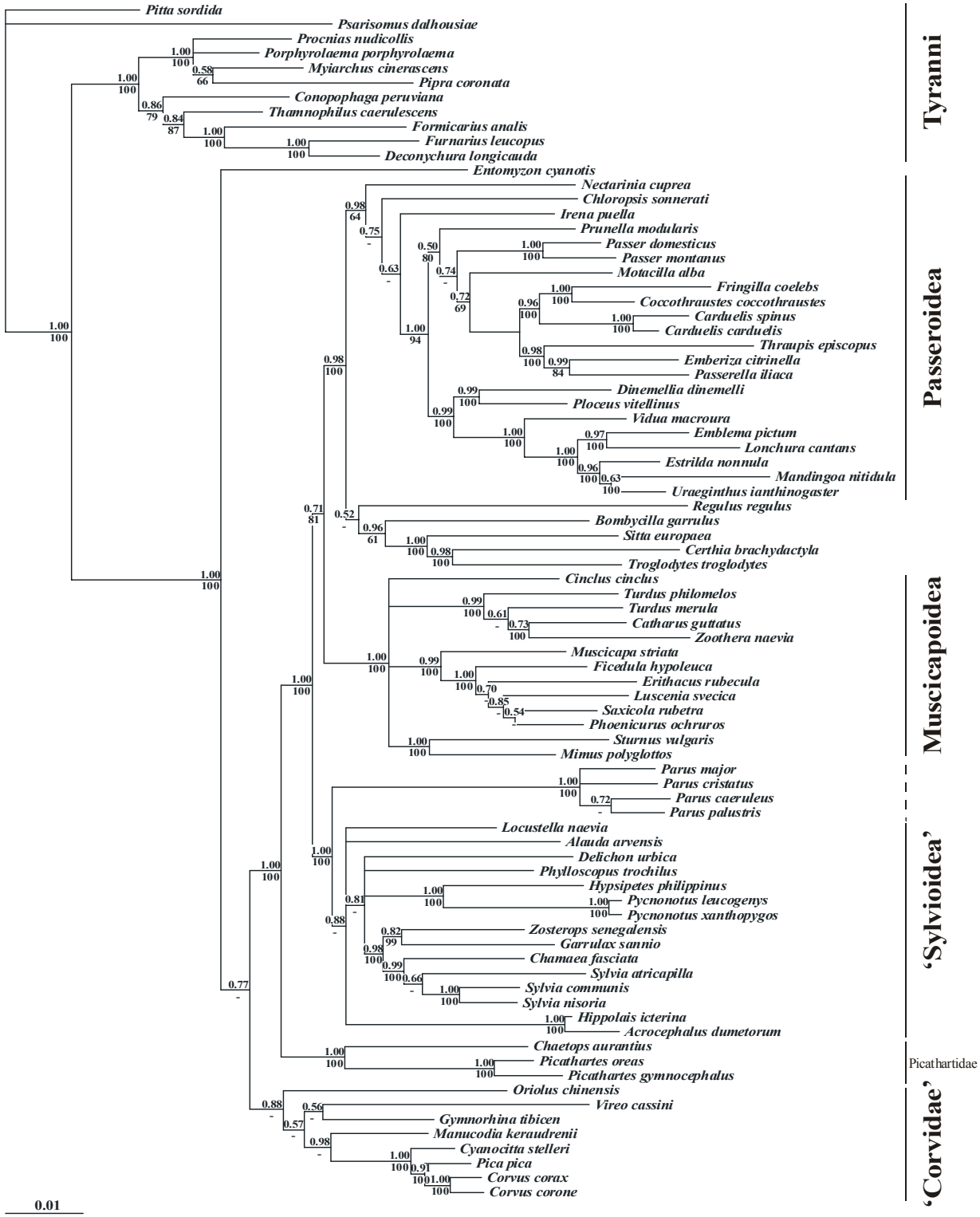


Fig. 1 Phylogenetic tree of the Bayesian analysis of the ZENK data set with Bayesian (upper value) and MP ratchet (lower value) support added at each node. Within Sylvioidea, the dashed line refers to 'Paroidea'.

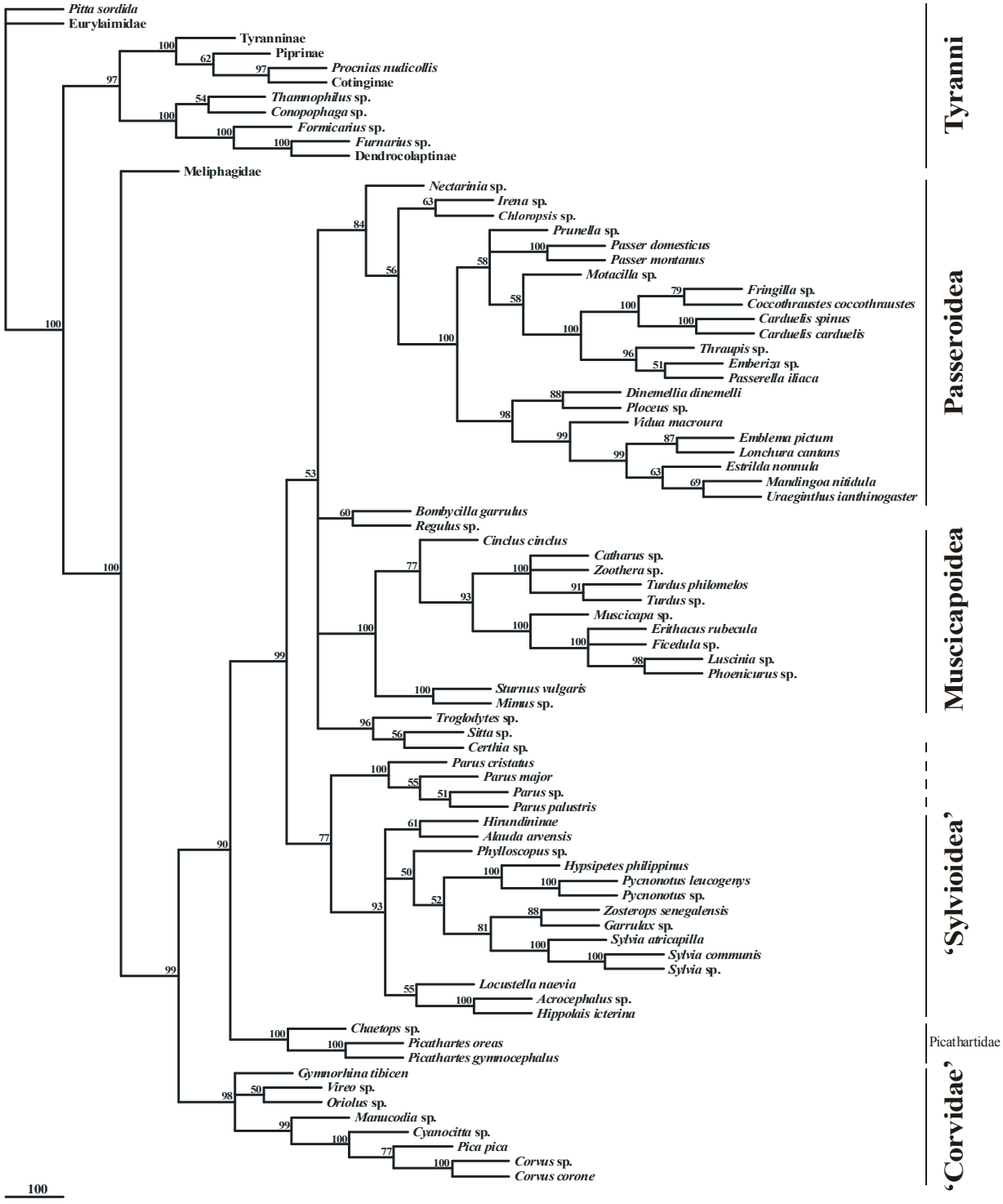


Fig. 2 Phylogenetic tree of the MP bootstrap analysis of the combined data set. Bootstrap support added at each node. When different representatives of a taxon originated sequences of the four genes, higher-level taxon names (i.e. genera or (sub)family) are given at the branches. Within Sylvioidea, the dashed line refers to 'Paroidea'.

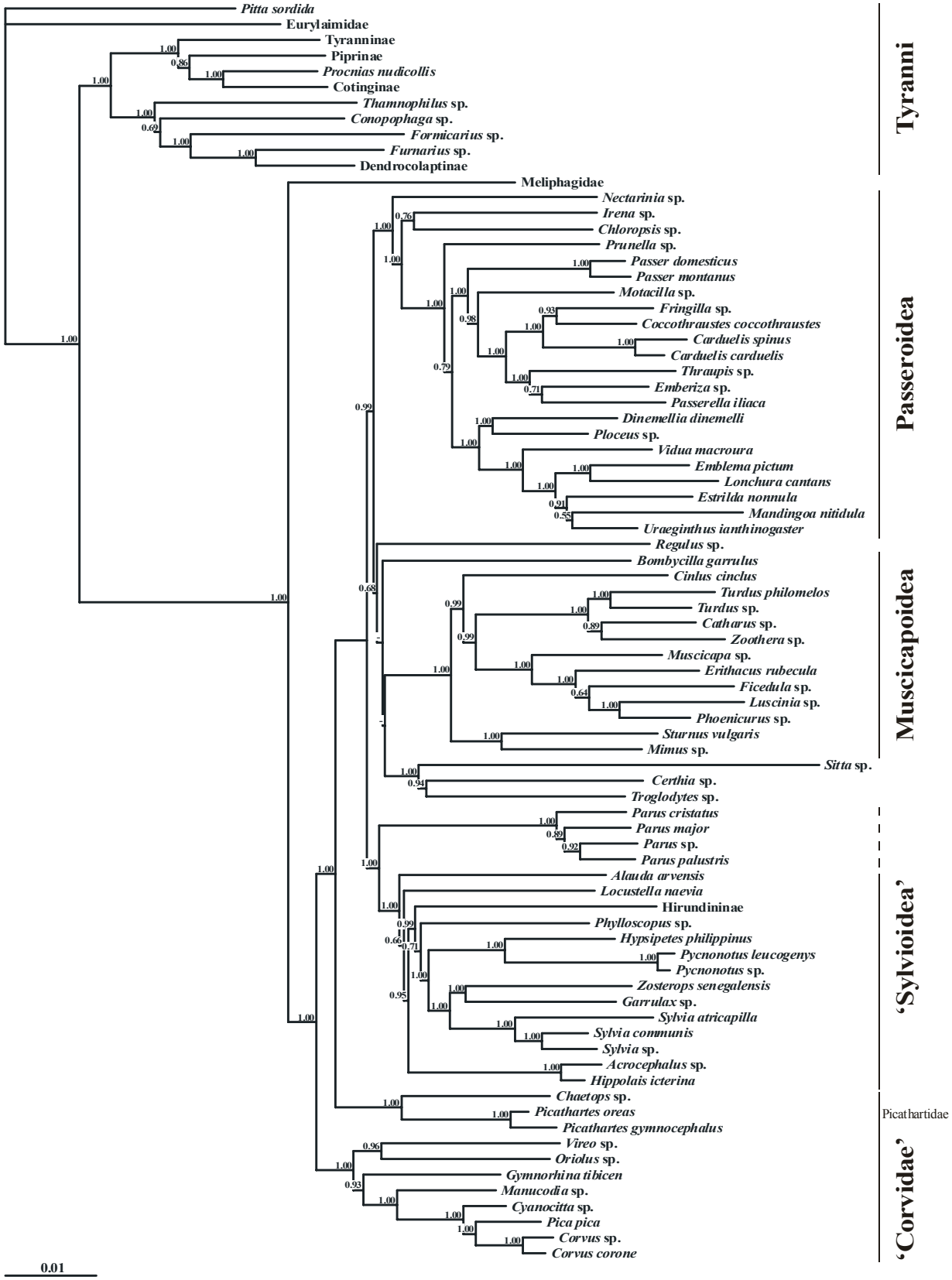


Fig. 3 Phylogenetic tree of the ML analysis of the combined data set with Bayesian support added at the nodes. When different representatives of a taxon originated sequences of the four genes, higher-level taxon names (i.e. genera or (sub)family) are given at the branches. Within Sylvioidea, the dashed line refers to 'Paroidea'.

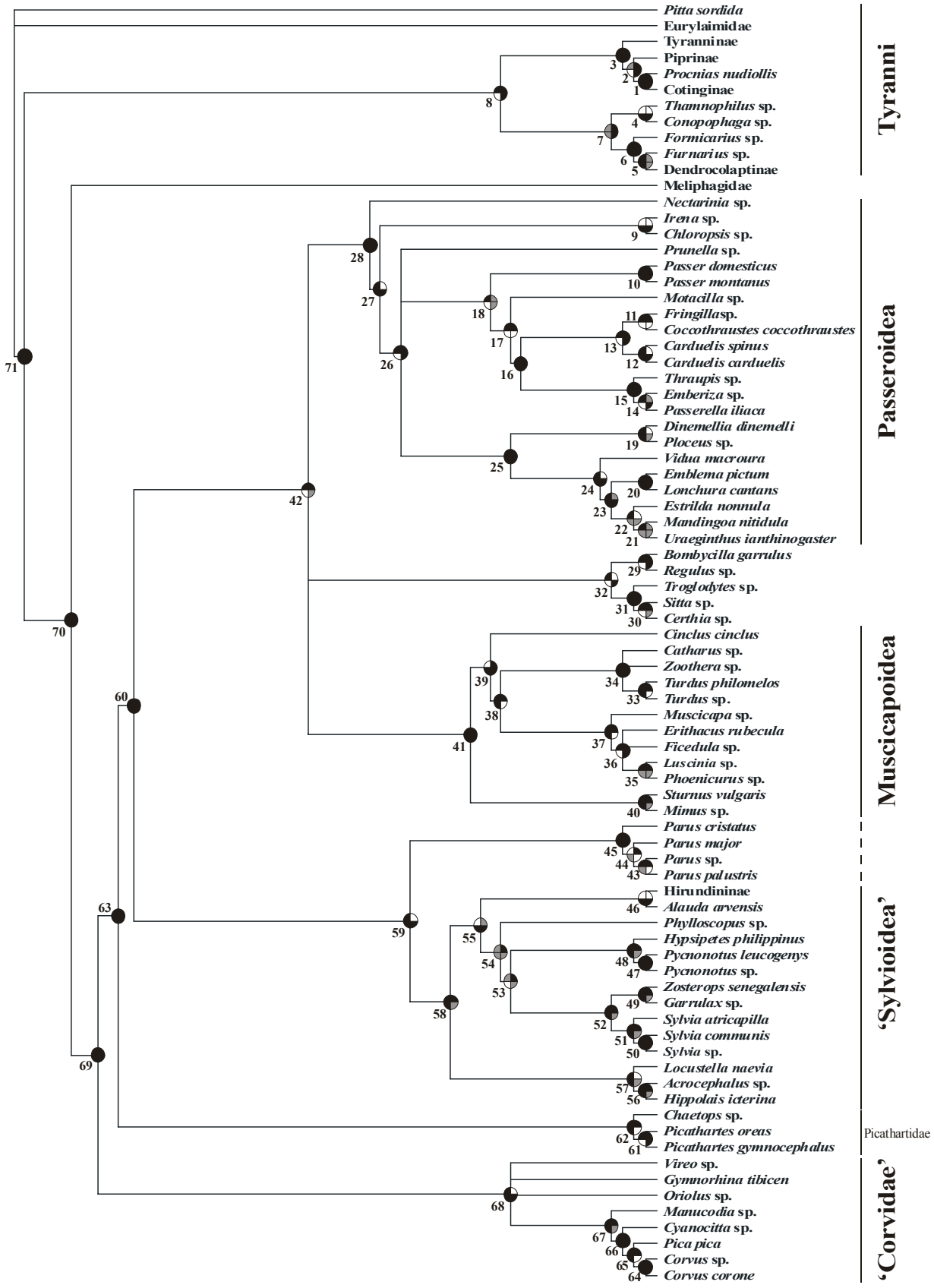


Fig. 4 Strict consensus tree of the MP ratchet analysis of the combined data set with Partitioned Bremer Support (PBS) added at each node. Black: positive PBS, grey: PBS=0, white: negative PBS. Quarters in circles refer to each gene as follows: upper left: ZENK, upper right: RAG-1, lower left: RAG-2, lower right: c-*myc*. When different representatives of a taxon originated sequences of the four genes, higher-level taxon names (i.e. genera or (sub)family) are given at the branches. Within Sylvioidea, the dashed line refers to 'Paroidea'.

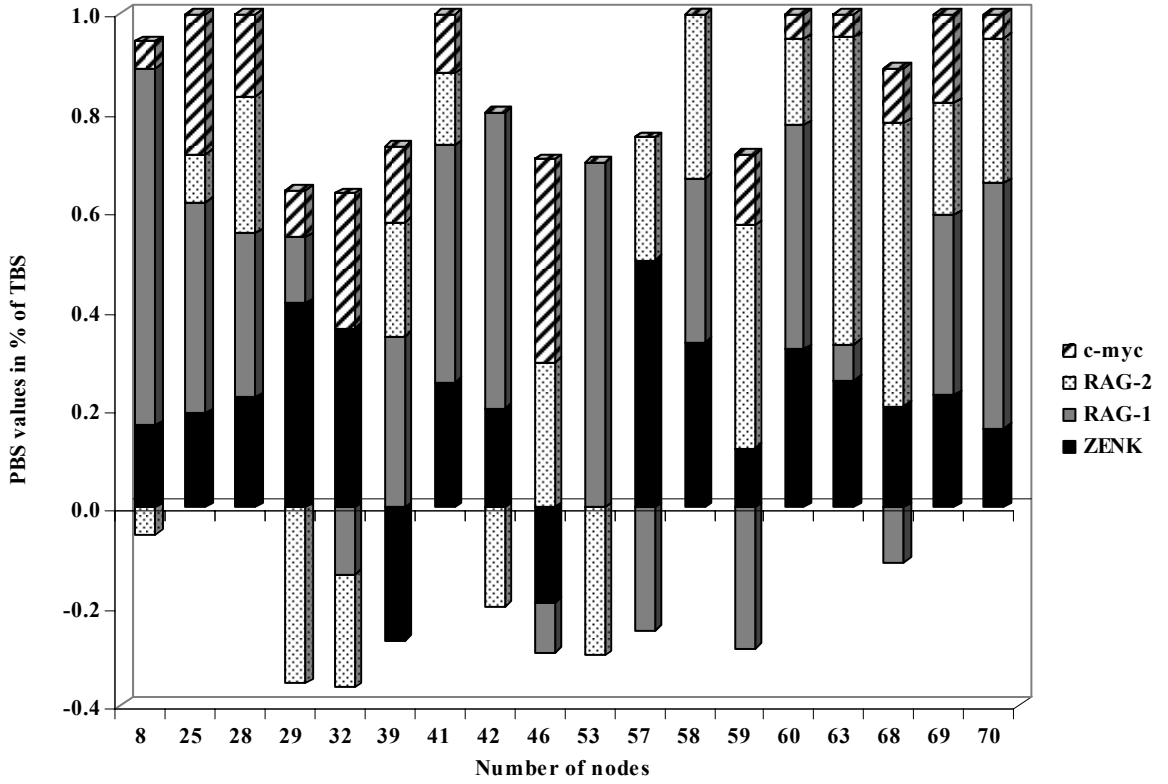


Fig. 5 Percentage PBS values of the four genes and their contribution to selected nodes. Number of nodes refers to Figure 4.

Supplementary Material

Table 1 Partitioned Bremer Support of each gene and total Bremer support, numbers of nodes refers to Figure 4 (article).

Number of node	ZENK	RAG-1	RAG-2	c-*myc*	total BS
1	0.50	1.50	2.50	2.50	7.00
2	0.00	3.00	-3.00	1.00	1.00
3	2.99	3.84	6.03	6.14	19.00
4	-1.00	-1.00	2.00	1.00	1.00
5	9.00	0.00	9.00	0.00	18.00
6	2.50	4.50	4.50	1.50	13.00
7	0.00	16.00	0.00	1.00	17.00
8	1.50	6.50	-0.50	0.50	8.00
9	-0.67	-2.00	4.67	1.00	3.00
10	10.33	21.00	9.17	3.50	44.00
11	3.67	1.00	-1.00	-0.67	3.00
12	12.50	-1.00	1.50	6.00	19.00
13	6.00	2.00	-0.50	0.50	8.00
14	1.00	0.00	-1.00	1.00	1.00
15	3.67	1.67	0.67	1.00	7.00
16	5.33	2.00	2.67	1.00	11.00
17	1.33	2.00	-0.33	-1.00	2.00
18	1.33	0.00	-0.33	0.00	1.00
19	3.50	-0.50	1.00	0.00	4.00
20	0.83	1.00	0.17	1.00	3.00
21	1.00	0.00	0.00	0.00	1.00
22	2.00	-1.00	0.00	0.00	1.00
23	5.00	0.00	2.50	1.50	9.00
24	8.00	-1.00	1.00	3.00	11.00
25	1.33	3.00	0.67	2.00	7.00
26	1.33	13.00	-2.33	1.00	13.00
27	0.46	-0.25	1.04	1.75	3.00
28	1.33	2.00	1.67	1.00	6.00
29	4.33	1.40	-3.73	1.00	3.00
30	1.00	1.67	-0.67	0.00	2.00
31	2.83	5.00	2.17	1.00	11.00
32	1.33	-0.50	-0.83	1.00	1.00
33	2.00	3.00	1.33	-1.33	5.00
34	11.50	24.50	6.33	0.67	43.00
35	0.33	7.67	0.00	0.00	8.00
36	3.00	6.00	-0.67	3.67	12.00
37	7.03	12.20	2.70	-1.93	20.00
38	0.97	5.00	1.18	-2.15	5.00
39	-1.17	1.50	1.00	0.67	2.00
40	4.33	4.00	9.67	0.00	18.00
41	6.33	12.00	3.67	3.00	25.00
42	0.33	1.00	-0.33	0.00	1.00
43	0.50	1.50	0.00	-1.00	1.00
44	-1.00	3.00	0.00	-1.00	1.00
45	35.00	17.00	7.00	3.00	62.00
46	-1.87	-1.00	2.87	4.00	4.00
47	22.00	11.00	13.00	9.00	55.00
48	8.00	12.00	3.00	0.00	23.00
49	1.67	3.67	1.67	0.00	7.00
50	2.00	4.00	1.00	2.00	9.00
51	2.00	9.00	4.00	0.00	15.00
52	2.00	1.00	1.00	0.00	4.00
53	0.00	3.50	-1.50	0.00	2.00
54	0.00	3.00	0.00	0.00	3.00
55	-1.00	0.00	1.00	1.00	1.00
56	28.00	6.00	7.00	0.00	41.00

57	2.00	-1.00	1.00	0.00	2.00
58	2.00	2.00	2.00	0.00	6.00
59	0.83	-2.00	3.17	1.00	3.00
60	3.52	5.00	1.92	0.56	11.00
61	8.96	19.75	-0.46	3.75	32.00
62	1.83	21.00	2.17	-1.00	24.00
63	1.55	0.43	3.74	0.29	6.00
64	3.00	4.00	5.00	2.00	14.00
65	1.00	1.00	1.00	-1.00	2.00
66	11.00	10.00	3.50	2.50	27.00
67	0.50	8.00	2.50	0.00	11.00
68	1.83	-1.00	5.17	1.00	7.00
69	2.50	4.00	2.50	2.00	11.00
70	12.83	39.67	23.50	4.00	80.00
71	2.00	11.00	3.50	4.50	21.00

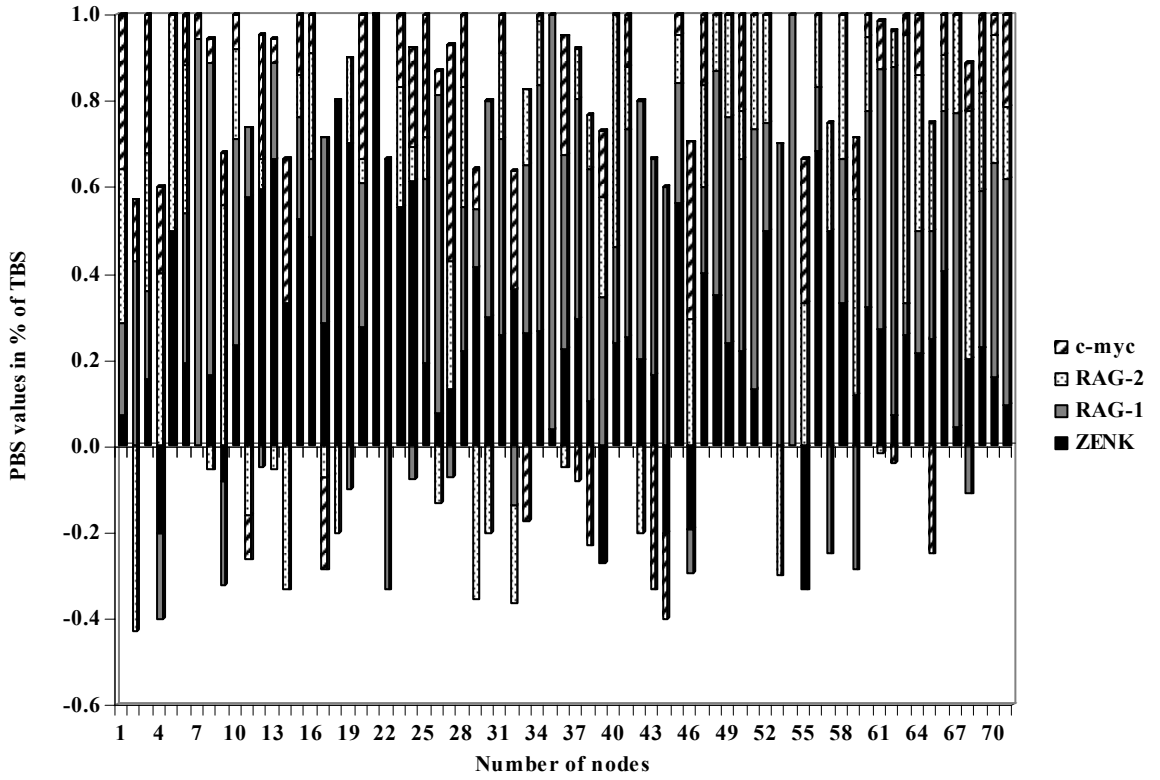


Fig. 1 Percentage PBS values of the four genes and their contribution to selected nodes, number of nodes refers to Figure 4 (article).

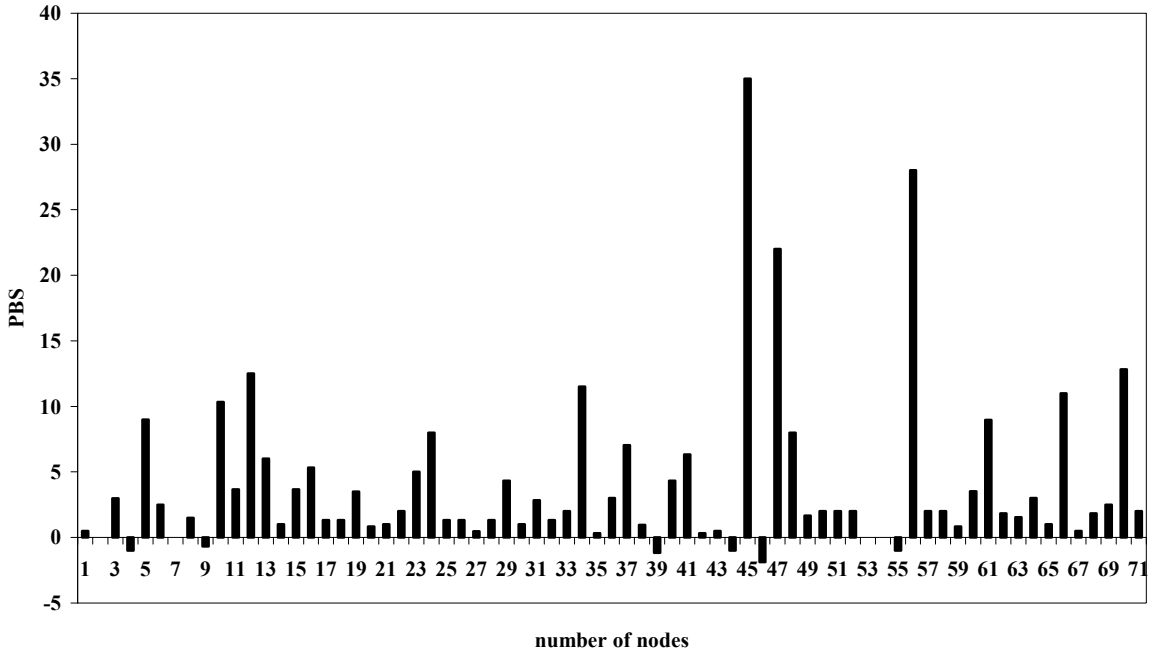


Fig. 2 Partitioned Bremer Support (PBS) values of the ZENK data set, number of nodes refers to Figure 4 (article).

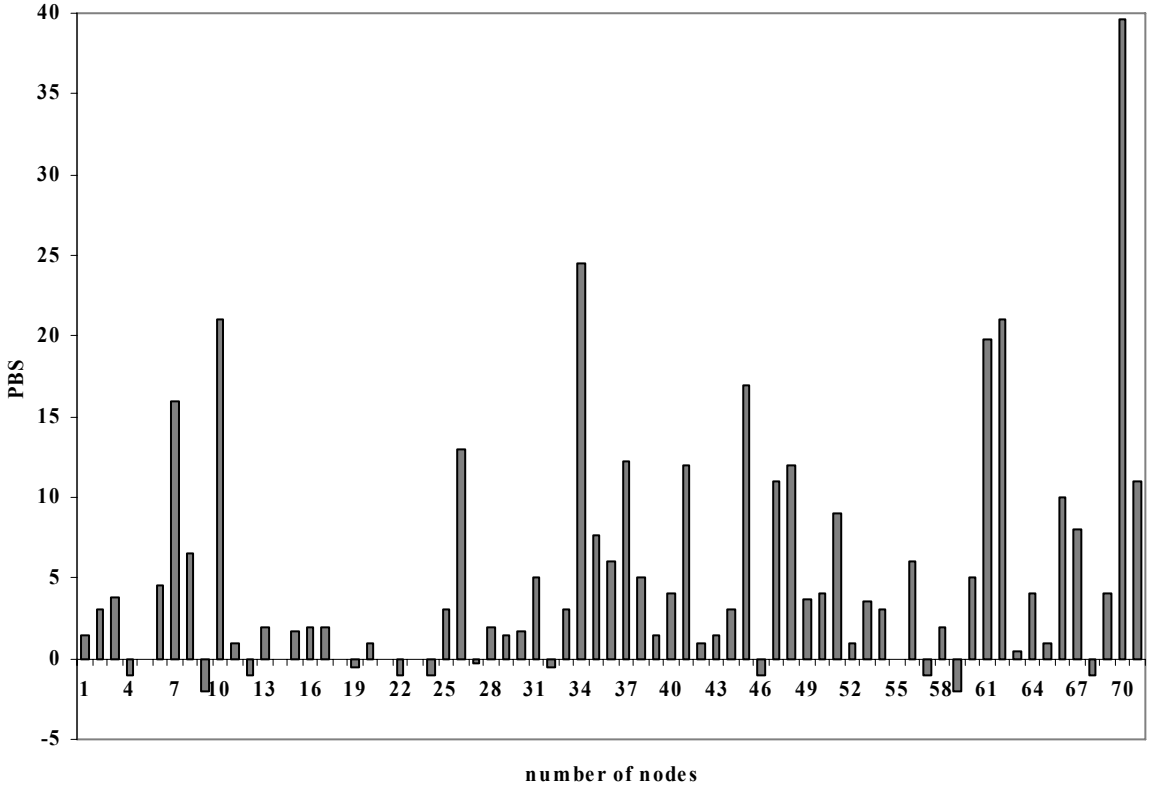


Fig. 3 Partitioned Bremer Support (PBS) values of the RAG-1 data set, number of nodes refers to Figure 4 (article).

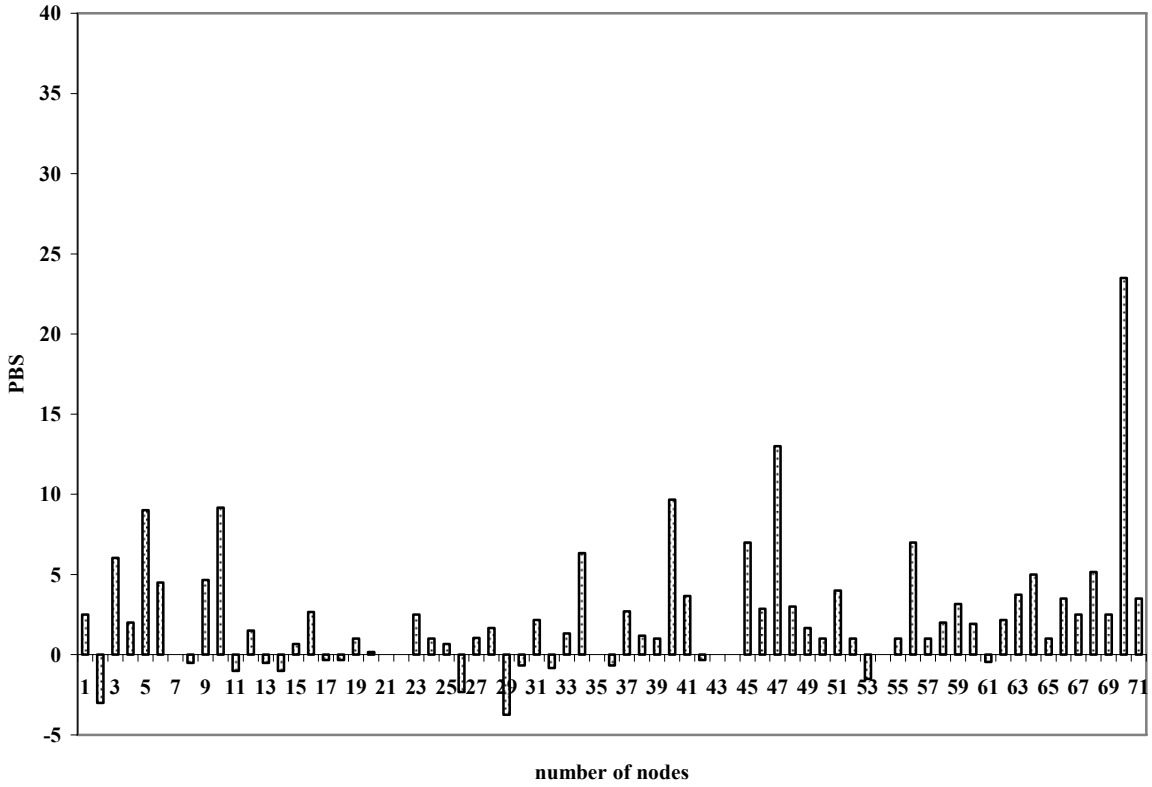


Fig. 4 Partitioned Bremer Support (PBS) values of the RAG-2 data set, number of nodes refers to Figure 4 (article).

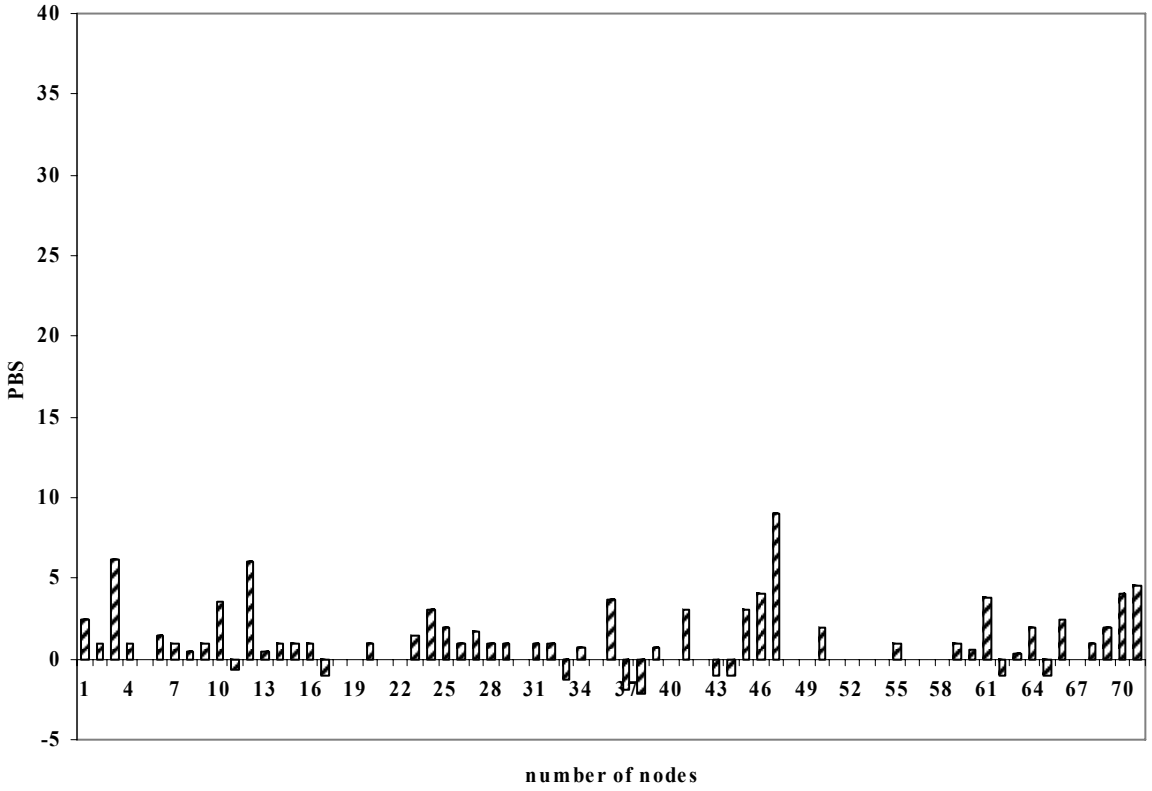


Fig. 5 Partitioned Bremer Support (PBS) values of the c-*myc* data set, number of nodes refers to Figure 4 (article).

7.2 Article II:

SIMONE TREPLIN and RALPH TIEDEMANN.

Specific chicken repeat 1 (CR1) retrotransposon insertion suggests phylogenetic affinity of rockfowls (genus *Picathartes*) to crows and ravens (Corvidae).

Molecular Phylogenetics and Evolution, under review.

Abstract

While the monophyly of the order Passeriformes as well as its suborders suboscines (Tyranni) and oscines (Passeri) is well established, both on morphological and molecular grounds, lower phylogenetic relationships have been a continuous matter of debate, especially within oscines. This is particularly true for the rockfowls (genus *Picathartes*), which phylogenetic classification has been an ongoing puzzle. Sequence-based molecular studies failed in deriving unambiguously resolved and supported hypotheses. We present here a novel approach: use of retrotransposon insertions as phylogenetic markers in passerine birds. Chicken repeat 1 (CR1) is the most important non-LTR retrotransposon in birds. We present two truncated CR1 loci in passerine birds, not only found in representatives of Corvinae (jays, crows and allies), but also in the West African *Picathartes* species which provide new evidence for a closer relationship of these species to Corvidae than has previously been thought. Additionally, we show that not only the absence/presence pattern of a CR1 insertion, but also the CR1 sequences themselves contain phylogenetic information.

Keywords: Picathartes, Corvidae, CR1, non-LTR retrotransposon, phylogeny

INTRODUCTION

The passerine birds (Passeriformes) are by far the largest avian order and, with more than 5700 species, comprise more than a half of all living birds (Sibley and Ahlquist, 1990). They form a morphologically very homogenous group and their monophyly is well established, both on morphological (Raikow, 1982) and molecular grounds (Sibley and Ahlquist, 1990). However, as most of the evolutionary lineages originated through a rapid radiation during the early Tertiary, phylogenetic relationships within the group have been a continuous matter of debate. The first extensive molecular study on avian systematics was based on DNA-DNA hybridization analyses (Sibley and Ahlquist (1990)). While subsequent sequence analyses generally corroborate partition of Passeriformes into two monophyletic clades, i.e., the suboscines (Tyranni) and the oscines (Passeri), conflicting phylogenetic hypotheses have been put forward for lower phylogenetic relationships, especially within oscines (e.g. Barker *et al.*, 2004; Beresford *et al.*, 2005; Ericson *et al.*, 2003; Ericson and Johansson, 2003).

Rockfowls, genus *Picathartes*, are endemic to the West African rainforest and consist of the two species grey-necked picathartes (*Picathartes oreas*) and white-necked picathartes (*Picathartes gymnocephalus*), which have long been regarded as avian curiosities (Thompson and Fotso, 1995). Due to their unique suit of morphological traits, the phylogenetic position of these species within Passeriformes has been the object of extensive debate and still remains a puzzle. *Picathartes gymnocephalus* was originally described as a crow (*Corvus gymnocephalus*, TEMMINCK 1825) before being assigned to its own genus *Picathartes* LESSON. Rockfowls were in turn placed within babblers (Amadon, 1943; Delacour and Amadon, 1951), starlings (Lowe, 1938), corvids (Sclater, 1930) and thrushes (Amadon, 1943). Sibley and Ahlquist (1990) supported the affinity to corvids and suggested the South African rockjumpers (genus *Chaetops*) as their closest relatives. *Chaetops* itself has been usually placed among babblers (McLachlan and Liversidge, 1978; Sclater, 1930; Sharpe, 1883) and thrushes (Swainson, 1832). Sibley and Ahlquist's (1990) analysis was inconclusive with regard to the phylogenetic affinity of the new family Picathartidae with both *Picathartes* and *Chaetops*, such that they assigned them to a separate parvorder with the status of *incertae sedis*, aside all other Passeri which were assembled into the two parvorders Corvida and Passerida. More recently, the sister relationship between *Picathartes* and *Chaetops* has been corroborated and Picathartidae has been regarded as sister and ancestral to Passerida (Barker *et al.*, 2004; Ericson and Johansson, 2003). Beresford *et al.* (2005) published a more extensive work with special emphasis on African endemic species. Their data supported the placement of Petroicidae as sister to Passerida (see also Barker *et al.*, 2004) and confirmed the basal position of the family Picathartidae within Passerida. They however suggested that a denser taxon sampling would have been desirable to resolve the deeper splits of the passerine phylogeny. Meanwhile, the validity of the major parvorders themselves has been questioned, as Corvida appear paraphyletic (Barker *et al.*, 2002; Ericson *et al.*, 2002a,b). Recent studies on passerine systematics differ in their taxonomic sampling, but generally rely on one or a few nuclear genes as phylogenetic markers, especially RAG-1, (Barker *et al.*, 2002; Barker *et al.*, 2004; Beresford *et al.*, 2005; Ericson and Johansson, 2003; Irestedt *et al.*, 2002; Irestedt *et al.*, 2001), RAG-2 (Barker *et al.*, 2004), c-*myc* (Ericson and Johansson, 2003; Ericson *et al.*, 2000; Irestedt *et al.*, 2002; Irestedt *et al.*, 2001) and myoglobin (Ericson and Johansson, 2003; Irestedt *et al.*, 2002). In the study presented here, we follow a different though complementary approach. We screened selected representative passerine taxa for occurrence of chicken repeat 1 (CR1) retrotransposon insertions (Stumph *et al.*, 1981). Retrotransposons are mobile genetic elements which are integrated in the genome via RNA intermediates. They can be

divided into a viral (containing retroviruses, long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons) and a nonviral superfamily (containing processed pseudogenes and short interspersed nucleotide elements (SINEs)) (Shedlock and Okada, 2000). SINE insertions are well established as molecular markers and have proved to be reliable apomorphic characters for phylogenetic inference (e.g., Huchon *et al.*, 2002; Lum *et al.*, 2000; Nikaido *et al.*, 2001; Nikaido *et al.*, 1999; Sasaki *et al.*, 2004; Schmitz *et al.*, 2001; Shedlock *et al.*, 2000; Shimamura *et al.*, 1997). With an estimated number of up to 100.000 copies in the chicken genome (Vandergon and Reitman, 1994), the chicken repeat 1 (CR1) is the most important non-LTR retrotransposon in birds. These elements consist of an 8bp direct repeat [typically (CATTCTRT) $(GATTCTRT)_{1-3}$ with some known variations] at the 3'-end, which can be easily detected (Silva and Burch, 1989). The vast majority of elements have severely truncated 5'-ends and have lost their retrotransposable ability (Silva and Burch, 1989; Stumph *et al.*, 1981). They are hence much shorter than the first complete consensus CR1 sequence published by Haas *et al.* (1997), which contained two complete open reading frames. The first study about the evolution of CR1 elements resulted in at least six different subfamilies (A-F); these results pointed to an ancient origin of these elements (Vandergon and Reitman, 1994). This hypothesis was confirmed and extended by finding CR1 elements in the genomes of other vertebrates (Chen *et al.*, 1991; Fantaccione *et al.*, 2004; Kajikawa *et al.*, 1997; Poulter *et al.*, 1999), while CR1 like elements have even been reported for some invertebrate species (Albalat *et al.*, 2003; Biedler and Tu, 2003; Drew and Brindley, 1997; Malik *et al.*, 1999). Despite their abundance in the avian genome and the clear character polarity of any single CR1 insertion at a particular locus, these non-LTR retrotransposons have so far only very rarely been used in a phylogenetic context. Recently, a single CR1 insertion in the lactate dehydrogenase B gene was used to support the monophyly of the Coscoroba-Cape Barren goose clade within Anseriformes (St. John *et al.*, 2005), while a second study considered the CR1 subfamily utility in the penguin phylogeny (Watanabe *et al.*, 2006). We here present hitherto undescribed CR1 elements, which are – together with our newly developed CR1 locus for the great tit (Treplin and Tiedemann, unpubl. results; GenBank accession no. **XXXXXX**) – the first CR1 described for passerine birds. The occurrence vs. absence of these elements at two distinct loci among selected passerine representatives will be used to build a phylogenetic framework for the placement of rockfowls as well as to shed light on affinities within Corvinae (jays, crows and allies *sensu* Sibley and Ahlquist (1990)).

MATERIALS AND METHODS

We amplified a partial CR1 element from the raven (*Corvus corax*) using the primers ParE10 114: 5'-TGGGCAGGGACACCTTCTACTAGACC-3' and Biotin-5'-GMMMMGGYTKCCCRRAGARGYTGTGG-3' (see GenBank accession no. **XXXXXX**). The CR1 identity of this amplificate was confirmed by comparison to different CR1 sequences from chicken (*Gallus gallus*) (Burch *et al.*, 1993; Haas *et al.*, 1997; Stumph *et al.*, 1983; Stumph *et al.*, 1981), sarus crane (*Grus antigone*) and emu (*Dromaius novaehollandiae*) (Chen *et al.*, 1991). We used this amplificate as a probe to establish a CR1 elements-enriched DNA genomic library from a total DNA extract of a raven liver sample, following a standard protocol for microsatellite enrichment (Paulus and Tiedemann, 2003). Recovered enriched fragments were transformed into competent *Escherichia coli* (TOPO cloning kit, Invitrogen). Recombinants were blotted onto a nylon membrane (Qiagen) and again hybridised with the CR1 probe. Positive clones were detected using the Phototope-Star chemiluminescent detection kit (New England Biolabs), sequenced with the BigDye version 3.1 Terminator Cycle Sequencing Kit (Applied Biosystems) and analysed on an AB 3100 multicapillary automatic sequencer (Applied Biosystems).

The cloned CR1 raven elements (Cor1-CR1 and Cor2-CR1, GenBank accession nos. **XXXXXX**) belong to the abundant 5'-truncated type of insertions, which complicates identification of the elements at the 5'-end. We used these sequences to design new specific primers. We located forward primers at the 5'-end of each clone, while reverse primers were located in the 3'-flanking region defined by the position of the CR1 specific 8 bp direct repeat identified in the clones (primers for the Cor1-CR1 locus: Cor1for 5'-GAGCAAACTTTATTCTATTATT-3', Cor1rev 5'-GCCATATTCTTTTGATTTCATT-3', and for the Cor2-CR1 locus: Cor2for 5'-GAATTCTTCCCTGTGAGG-3', Cor2rev 5'-GCCGTTTTGCTGCTTACCATA-3'). These primers yielded single amplicons (335 bp for Cor1-CR1; 251 bp for Cor2-CR1) from raven genomic DNA. They were also used for PCR-amplifications on genomic DNA from different passerine birds supposed to be closely related to the raven (following Sibley and Monroe (1990)) (Tab. 1). PCR-amplifications were performed in a total volume of 37.5 µl, containing 1 mM Tris-HCl, pH 9.0, 5 mM KCl, 0.15 mM $MgCl_2$, 0.05 mM of each dNTP, 0.13 µM of both forward and reverse primers and 0.75 U *Taq* polymerase (Qbiogene) in a Biometra T3000 thermocycler according to the following reaction profile: 1 cycle at 96°C for 5 min, 40 cycles at 96°C for 1 min 30 s, 54°C and 55°C (for the Cor1-CR1 and Cor2-CR1 respectively) for 1 min 15 s, 72°C for 1 min 30 s, and a

final extension at 72°C for 10 min. Cycle sequencing reactions were performed using the forward and reverse primers.

Sequences were aligned in the BioEdit Sequence Alignment Editor (Hall, 1999) and analysed phylogenetically by maximum parsimony (MP) (Farris *et al.*, 1970) using the heuristic search option with the TBR-branch-swapping algorithm in PAUP* 4.0b10 (Swofford, 2001). Robustness of the phylogenetic hypotheses were evaluated by bootstrapping (Felsenstein, 1985) with 1000 replicates. We used MrModeltest version 2.2 (Nylander, 2004) to identify the best model of sequence evolution for both datasets. Data were analysed under a maximum likelihood (ML) criterion in a Bayesian framework (Huelsenbeck *et al.*, 2000; Larget and Simon, 1999; Mau and Newton, 1997; Mau *et al.*, 1999; Rannala and Yang, 1996) using MrBayes 3.0b4 (Huelsenbeck and Ronquist, 2001), with the two models yielded (GTR for Cor1-CR1 and K80+I for Cor2-CR1) and each analysis starting from a random tree. The program was set to run 10^6 generations with four Markov Chain Monte Carlo iterations simultaneously and trees sampled every 100 generations with the first 200 of each discarded, as estimated graphically as burn-in.

For taxa where PCR-amplifications of our CR1 elements failed, we performed hybridization experiments to evaluate the presence/absence of the respective loci. As hybridization to single CR1 loci might not yield a detectable signal in Southern blots, especially with non-radioactive methods (St. John *et al.*, 2005), we applied the more sensitive dot blot technique here. This analysis was performed on those samples where a sufficient amount of high quality DNA (large molecular size) was available. 5 µg of genomic DNA from carrion crow (*Corvus corone*), Bohemian waxwing (*Bombycilla garrulus*), black redstart (*Phoenicurus ochruros*) and great tit (*Parus major*) were dot-blotted onto a nylon membrane (Qiagen) and hybridised with a biotin-labelled probe, spanning over the insertion site at the 3'-end of the element into the flanking region (designed to bind specifically to the Cor2 locus, cf. Fig. 3) overnight at 65°C. Membranes were washed with increasing stringency (0.5 x SSC / 0.1% SDS at room temperature, 0.5 x SSC / 0.1% SDS at 45°C, 0.1x SSC / 0.1% SDS at 65°C).

RESULTS

In the raven CR1 enriched genomic library we were able to identify two CR1 elements by direct comparisons to the complete CR1 consensus sequence described by Haas *et al.* (1997).

We sequenced 372 bp of the locus Cor1, containing the 3'-end of reverse transcriptase and the CR1 direct repeat (Fig. 1). A blastn search in the NCBI database yielded a perfect match of 35 bp within a CR1 locus of the grey petrel (*Procellaria cinerea*). At locus Cor2 we sequenced a 283 bp long fragment. Both the 3'-end of reverse transcriptase and the direct repeat were again detectable (Fig. 1). A blastn search yielded an 89% match of 144 bp with a CR1 gene for the chicken repeat 1 of the king penguin (*Aptenodytes patagonicus*). Translation into amino acid sequences and protein/protein search yielded similarities with reverse transcriptase of the second open reading frame (ORF2) of the chicken for Cor1 and Cor2. The 5'-end of a CR1 can be often (but not always, Vandergon and Reitman, 1994)) identified by a 6 bp target site duplication directly adjacent to the inserted CR1 element (Silva and Burch, 1989). We did not detect such duplication for our CR1 loci, which might be either due to mutations after the insertion event (obscuring the duplication pattern) or to a lack of the 5'-end in the sequences we analysed. Therefore, we performed our locus-specific PCRs using a primer adjacent to the 3' end of the CR1 together with one element-specific internal primer.

We confirmed the presence of the Cor1-CR1 locus by PCR-amplifications and subsequent sequencing in six Corvinae species, representing all tribes except Oriolini; the representative of the latter tribe (*Oriolus chinensis)* did not give successful amplifications (Table 1). We were able to amplify and sequence this locus only for taxa belonging to the subfamily Corvinae; we therefore conclude that, given our taxon sampling, the presence of the Cor1-CR1 locus is restricted to this subfamily.

We confirmed the presence of the Cor2-CR1 locus by PCR-amplifications and subsequent sequencing in all seven tested corvine (including *Oriolus chinensis*) and in the two *Picathartes* species (Table 1). Successful amplifications of the locus were limited to Corvinae and Picathartidae (Table 1). Two additional taxa (*Bombycilla garrulus* and *Cinclus cinclus*), however, yielded PCR-amplificates 264 bp and 350 bp long, respectively. These amplificates differed in size from the Cor2-CR1 sequences (283 bp long) we obtained for Corvinae and Picathartidae. Sequencing of the fragments revealed two CR1 type elements, which showed some similarity to our Cor2-CR1 locus (Fig. 3). However, sequence analyses, both of the 3'-end of the element indicated by the direct repeat (cf. Fig 3) and the downstream locus-specific

sequence, showed pronounced divergence in sequence and length in *Bombycilla* and *Cinclus*, compared to the other taxa. We hence conclude that these two PCR-amplificates represent related, but non-orthologous CR1 elements, i.e., independent insertion events at different loci (named Cor2-CR1 *like* elements hereafter).

To test whether the failures in the PCR amplification of the Cor2-CR1 locus might be due to a real absence of the locus in some of the species we included in the study or to technical problems (i.e., nucleotide substitutions in the primer binding sites), we performed a locus-specific dot blot on the genomic DNA. Hybridization to our specific Cor2-CR1 probe was tested in dot blotting (a) the Cor2-CR1 PCR products (Fig. 4, dot 1-5), (b) the Cor1-CR1 PCR product (dot 6 and 7), and water (dot 8) as a negative control and genomic DNA of several representatives of the Passeri (dot 9-12). As expected, the probe hybridised well with the genomic DNA of the carrion crow (*Corvus corone*, dot 9) and negligibly to the black redstart (*Phoenicurus ochruros*, dot 11) and the great tit (*Parus major*, dot 12). Additionally, the probe also hybridised with the genomic DNA of the Bohemian waxwing (*Bombycilla garrulus*, dot 10). Based on the sequence comparison of the PCR amplicon produced by the Cor2-CR1 in *Bombycilla garrulus*, we consider the hybridizing locus in this species a Cor2-CR1 *like* locus (see above, cf. Fig. 3), non-orthologous to our Cor2-CR1 locus.

Sequencing of the Cor1-CR1 locus revealed three single nucleotide indels (Fig. 2), whereas the Cor2-CR1 locus contained some indels comprising more than one nucleotide (Fig. 3). A six basepair deletion was found in the sequence of the carrion crow (*Corvus corone*). The magpie (*Pica pica*) and Steller's jay (*Cyanocitta stelleri*) shared one insertion of two basepairs and the white-backed magpie (*Gymnorhina tibicen*) showed a deletion of sixteen basepairs. Sequences of the two rockfowl species (genus *Picathartes*) were identical.

Phylogenetic analyses of the Corvinae-specific Cor1-CR1 locus support the monophyly of the genus *Corvus* and the placement of *Corvus* and *Pica* as each other's closest relatives (Fig. 5). Phylogenetic analyses of the Cor2-CR1 locus (Fig. 6) recovered, additionally to the corroborated results of the Cor1-CR1 analysis (Fig. 5), both Corvini and Picathartidae as monophyletic groups but, due to the obvious lack of a clear outgroup, the relationships among the different tribes/family included in the tree cannot be safely assessed, as they appear almost equally distant from one another.

DISCUSSION

Phylogenetic information contained in CR1 insertions

The wide distribution and high conservation of chicken repeat 1 retrotransposons in birds was first shown by Chen *et al.* (1991) and led to a copious characterization of these elements in the chicken genome. Based on the properties of the conservative 3'-end region, suggested to act as a recognition site for reverse transcriptase (Eickbush, 1992), we were able to characterise phylogenetically informative CR1 elements for passerine birds.

The CR1 loci we cloned from the raven genome showed high similarities with the chicken complete consensus CR1 (Haas *et al.* 1997) in the region of the second open reading frame that codes for a reverse transcriptase (Burch *et al.*, 1993). The 5'-end of the elements could not be unambiguously detected. This was not unexpected, as CR1 elements are frequently truncated and the only definite indication of the 5'-end is a 6 bp target site duplication, which is often lacking (Vandergon and Reitman, 1994). Therefore, we designed specific primers away from the 5'-end to avoid this problem (see Materials and Methods). Evidently with such an approach, false-negative PCR results (lack of amplifications due to substitution at the primer site) could not be completely ruled out. Nevertheless, we consider our loci phylogenetically informative, based on two lines of reasoning: (1) False negative results do not question the phylogenetic relationships of those species for which we obtained positive PCR amplifications (Shedlock and Okada, 2000), (2) in case of non-amplification we confirmed absence of the CR1 element by a hybridization experiment. As the presence/absence pattern was fully consistent among the locus-specific hybridization on genomic DNA and the PCR amplification, we conclude that those taxa where our analyses did not indicate presence of Cor1-CR1 and/or Cor2-CR1 elements, truly lack those elements at the considered loci. Based on this argument and the mode of retrotransposon insertion (see above), we consider presence of our new CR1 loci – based on our combined PCR/hybridization evidence – an apomorphic character state.

Phylogeny of Corvidae and Picathartidae

If we assume the presence of a CR1 locus as an apomorphic character state then (1) Picathartidae and Corvinae form a monophyletic clade because they all share the Cor2-CR1 locus and (2) the presence of Cor1-CR1 groups Corvini, Artamini, and Paradisaeini in a monophyletic cluster (Fig. 7).

Within Corvinae we detected the locus Cor1-CR1 in all representatives of Corvini, Paradisaeini, and Artamini we tested, supporting the monophyly of these three groups, but not in the Oriolini (Figs. 2, 7). This locus therefore provides additional evidences on the phylogeny of the subfamily Corvinae, pointing to a close relationship among Corvini, Paradisaeini and Artamini. A sister taxa relationship between birds of paradise (Paradisaeini) and corvids (Corvini), has been repeatedly postulated (Cracraft and Feinstein, 2000; Frith and Beehler, 1998; Helmbychowski and Cracraft, 1993; Nunn and Cracraft, 1996), as well as the hypothesis of these three tribes being a monophyletic group within Corvinae (Helmbychowski and Cracraft, 1993). Sequence-based studies confirmed a close affinity between Corvini and Paradisaeini but, at the same time, were not able to provide definitive evidence on the phylogenetic positions of Artamini and Oriolini as well as on the monophyly of Corvinae (*sensu* Sibley and Ahlquist, 1990) (Barker *et al.*, 2002; Barker *et al.*, 2004; Beresford *et al.*, 2005). Given our limited taxonomic sampling of Corvidae (*sensu* Sibley and Ahlquist, 1990), we are presently unable to contribute further to this point. However, the fact that orioles (conventionally included in Corvinae) possess only the Cor2-CR1 locus is an interesting new aspect to the phylogeny within Corvinae.

Our hypothesis of *Picathartes* being closer to corvids than to Passerida (Fig. 7) is at odds with many molecular studies. Sibley and Ahlquist (1990) tentatively concluded that *Picathartes* should have affinities to Corvida, but they conveyed their uncertainty, coupled with ambiguous morphological data, by placing the genus between Corvida and Passerida with the status of *incertae sedis*. The first sequence-based study including *Picathartes* was not able to resolve its phylogenetic position (Barker *et al.*, 2002). Ericson and Johansson (2003) proposed *Picathartes* and *Chaetops* being basal to Passerida. They classified them as Passerida because the three groups all share a 3 bp insertion in the sequence of the proto-oncogene c-*myc* (exon 3), a character considered apomorphic for Passerida. The position of Picathartidae as the earliest branch of Passerida has been further proposed by Barker *et al.* (2004) and Beresford *et al.* (2005). However, while the monophyly of Passerida (excluding Picathartidae) is apparently well established, the placement of *Picathartes* within this group never gained a robust statistical support. Additionally, Beresford *et al.* (2005), by proposing Petroicidae as the second branch in Passerida (branching off after Picathartidae), challenged the assumption of the 3 bp insertion in c-*myc* being an apomorphy for Passerida, as Petroicidae lacks this insertion (only available representative *Eopsaltria australis* (Ericson *et al.*, 2002b)). Recently, Fuchs *et al.* (2006) and Jønsson and Fjeldså (2006) highlighted the difficulties in recovering a robust phylogenetic hypothesis at the boundary between "Corvida"

and Passerida using sequence data. These contrasting results suggest that the character polarity of the 3 bp length difference, i.e., whether it comprises an insertion or a deletion in the c-*myc* gene, is presumably difficult to assess. In contrast, for a retrotransposition event, as reported in our study, the character state is clear-cut, i.e., apomorphic, and homoplasy, i.e., independent insertion events of the same element at the same locus, very unlikely. Hence, we conclude that our CR1 data strongly support a phylogenetic hypothesis, which places Picathartidae closer to Corvidae than to Passerida.

If we assume that the presence of a CR1 element at a particular locus in a given group of species is the result of a single insertion event, which occurred in their common ancestor's genome, such a locus should also be phylogenetically informative on the relationships within the group itself. Our data support this assumption, as the trees based on Cor1 and 2 sequences both showed high resolution in defining genera and tribes (Figs. 5 and 6). Additionally, they are in agreement with well-accepted phylogenetic hypothesis, i.e., the monophyly of the genus *Corvus* (crows and ravens) and its closer relationship to *Pica pica* than to *Cyanocitta stelleri* (Cibois and Pasquet, 1999; Ericson *et al.*, 2005).

One might challenge our interpretation that the Cor2-CR1 *like* locus consistently detected by both amplification and hybridization in *Bombycilla* is non-orthologous to our Cor2-CR1 locus, found in Picathartidae and Corvinae. In fact, this statement on non-orthology is based on the pronounced sequence difference at the 3' end among the *Bombycilla* sequence and all sequences of Picathartidae and Corvidae (cf. Fig. 3). We have also considered the alternative interpretation that *Bombycilla* also possesses the orthologous Cor2-CR1 locus. While this alternative argument would not alter any of our phylogenetic hypotheses about Corvinae and their relationship to Picathartidae, it would additionally point towards a phylogenetic affinity of waxwings (*Bombycilla*) to a combined Picathartidae/Corvidae clade. While such a relationship would evidently comprise an interesting additional phylogenetic hypothesis to be further evaluated, we prefer to consider our data on *Bombycilla* inconclusive so far.

ACKNOWLEDGEMENTS

We thank Hazell Thompson and Roger Fotso for providing *Picathartes* samples, the Zoological Garden Berlin for providing blood samples of *Manucodia*, *Gymnorhina* and Irenidae, Joachim Oesert and Heiner Luttmann for providing tissue samples of the Bohemian waxwing, dipper, and robin, Malte Mews for providing tissue samples of *Cyanocitta*, and Tomek Janiszewski for a sample of *Pica pica*. Christoph Bleidorn and Valerio Ketmaier provided valuable comments on the manuscript.

REFERENCES

Albalat, R., Permanyer, J., Martinez-Mir, C., Gonzalez-Angulo, O., Gonzalez-Duarte, R., 2003. The first non-LTR retrotransposon characterised in the cephalochordate amphioxus, BfCR1, shows similarities to CR1 like elements. Cell. Mol. Life Sci. 60, 803-809.

Amadon, D., 1943. The genera of starlings and their relationships. Am. Mus. Novit. 1247, 1-16.

Barker, F. K., Barrowclough, G. F., Groth, J. G., 2002. A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. Proc. R. Soc. Lond. B Biol. Sci. 269, 295-308.

Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., Cracraft, J., 2004. Phylogeny and diversification of the largest avian radiation. Proc. Natl. Acad. Sci. U.S.A. 101, 11040-11045.

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 Biol. Sci. 272, 849-858.

Biedler, J., Tu, Z. J., 2003. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: Unprecedented diversity and evidence of recent activity. Mol. Biol. Evol. 20, 1811-1825.

Burch, J. B. E., Davis, D. L., Haas, N. B., 1993. Chicken repeat-1 elements contain a *pol*-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. Proc. Natl. Acad. Sci. U.S.A. 90, 8199-8203.

Chen, Z. Q., Ritzel, R. G., Lin, C. C., Hodgetts, R. B., 1991. Sequence conservation in avian CR-1 - an interspersed repetitive DNA family evolving under functional constraints. Proc. Natl. Acad. Sci. U.S.A. 88, 5814-5818.

Cibois, A., Pasquet, E., 1999. Molecular analysis of the phylogeny of 11 genera of the Corvidae. Ibis 141, 297-306.

Cracraft, J., Feinstein, J., 2000. What is not a bird of paradise? Molecular and morphological evidence places *Macgregoria* in the Meliphagidae and the Cnemophilinae near the base of the corvoid tree. Proc. R. Soc. Lond. B Biol. Sci. 267, 233-241.

Delacour, J., Amadon, D., 1951. The systematic position of *Picathartes*. Ibis 93, 60-62.

Drew, A. C., Brindley, P. J., 1997. A retrotransposon of the non-long terminal repeat class from the human blood fluke *Schistosoma mansoni*. Similarities to the chicken-repeat-1-like elements of vertebrates. Mol. Biol. Evol. 14, 602-610.

Eickbush, T. H., 1992. Transposing without ends - the non-LTR retrotransposable elements. New Biologist 4, 430-440.

Ericson, P. G. P., Johansson, U. S., Parsons, T. J., 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, P. G. P., Christidis, L., Cooper, A., Irestedt, M., Jackson, J., Johansson, U. S., Norman, J. A., 2002a. A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. Proc. R. Soc. Lond. B Biol. Sci. 269, 235-241.

Ericson, P. G. P., Christidis, L., Irestedt, M., Norman, J. A., 2002b. Systematic affinities of the lyrebirds (Passeriformes: *Menura*), with a novel classification of the major groups of passerine birds. Mol. Phylogenet. Evol. 25, 53-62.

Ericson, P. G. P., Irestedt, M., Johansson, U. S., 2003. Evolution, biogeography, and patterns of diversification in passerine birds. J. Avian Biol. 34, 3-15.

Ericson, P. G. P., Johansson, U. S., 2003. Phylogeny of Passerida (Aves: Passeriformes) based on nuclear and mitochondrial sequence data. Mol. Phylogenet. Evol. 29, 126-138.

Ericson, P. G. P., Jansen, A. L., Johansson, U. S., Ekman, J., 2005. Inter-generic relationships of the crows, jays, magpies and allied groups (Aves: Corvidae) based on nucleotide sequence data. J. Avian Biol. 36, 222-234.

Fantaccione, S., Russo, C., Palomba, P., Rienzo, M., Pontecorvo, G., 2004. A new pair of CR1 like LINE and tRNA-derived SINE elements in *Podarcis sicula* genome. Gene 339, 189-198.

Farris, J. S., Kluge, A. G., Eckardt, M. J., 1970. A numerical approach to phylogenetic systematics. Syst. Zool. 19, 172-189.

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

Frith, C. B., Beehler, B. M., 1998. The birds of paradise. Oxford University Press, New York, NY.

Fuchs, J., Fjeldså, J., Bowie, R. C., 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.

Haas, N. B., Grabowski, J. M., Sivitz, A. B., Burch, J. B. E., 1997. Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames. Gene 197, 305-309.

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

Helmbychowski, K., Cracraft, J., 1993. Recovering phylogenetic signal from DNA-sequences - relationships within the corvine assemblage (Class Aves) as inferred from complete sequences of the mitochondrial-DNA cytochrome-*b* gene. Mol. Biol. Evol. 10, 1196-1214.

Huchon, D., Nikaido, M., Okada, N., 2002. Use of SINE retroposon in phylogeny. M S-Medecine Sciences 18, 1276-1281.

Huelsenbeck, J. P., Rannala, B., Masly, J. P., 2000. Accommodating phylogenetic uncertainty in evolutionary studies. Science 288, 2349-2350.

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

Irestedt, M., Johansson, U. S., Parsons, T. J., Ericson, P. G. P., 2001. Phylogeny of major lineages of suboscines (Passeriformes) analysed by nuclear DNA sequence data. J. Avian Biol. 32, 15-25.

Irestedt, M., Fjeldså, J., Johansson, U. S., Ericson, P. G. P., 2002. Systematic relationships and biogeography of the tracheophone suboscines (Aves: Passeriformes). Mol. Phylogenet. Evol. 23, 499-512.

Jønsson, K. A., Fjeldså, J., 2006. A phylogenetic supertree of oscine passerine birds (Aves: Passeri). Zool. Scr. 35, 149-186.

Kajikawa, M., Ohshima, K., Okada, N., 1997. Determination of the entire sequence of turtle CR1: The first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. Mol. Biol. Evol. 14, 1206-1217.

Larget, B., Simon, D. L., 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16, 750-759.

Lowe, P. R., 1938. Some anatomical and other notes on the systematic position of the genus *Picathartes*, together with some remarks on the families Sturnidae and Eulabetidae. Ibis, 254-269.

Lum, J. K., Nikaido, M., Shimamura, M., Shimodaira, H., Shedlock, A. M., Okada, N., Hasegawa, M., 2000. Consistency of SINE insertion topology and flanking sequence tree: quantifying relationships among cetartiodactyls. Mol. Biol. Evol. 17, 1417-1424.

Malik, H. S., Burke, W. D., Eickbush, T. H., 1999. The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. 16, 793-805.

Mau, B., Newton, M. A., 1997. Phylogenetic inference for binary data on dendograms using Markov chain Monte Carlo. J. Comput. Graph. Stat. 6, 122-131.

Mau, B., Newton, M. A., Larget, B., 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Biometrics 55, 1-12.

McLachlan, G. R., Liversidge, R., 1978. Roberts birds of South Africa, 3rd ed. Trustees of the John Voelcker bird book fund, Cape Town.

Nikaido, M., Rooney, A. P., Okada, N., 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. U.S.A. 96, 10261-10266.

Nikaido, M., Matsuno, F., Hamilton, H., Brownell, R. L., Cao, Y., Ding, W., Zuoyan, Z., Shedlock, A. M., Fordyce, R. E., Hasegawa, M., Okada, N., 2001. Retroposon analysis of major cetacean lineages: The monophyly of toothed whales and the paraphyly of river dolphins. Proc. Natl. Acad. Sci. U.S.A. 98, 7384-7389.

Nunn, G. B., Cracraft, J., 1996. Phylogenetic relationships among the major lineages of the birds-of-paradise (Paradisaeidae) using mitochondrial DNA gene sequences. Mol. Phylogenet. Evol. 5, 445-459.

Nylander, J. A. A., 2004. MrModeltest v2. Program distributed by the author.

Paulus, K. B., Tiedemann, R., 2003. Ten polymorphic autosomal microsatellite loci for the Eider duck *Somateria mollissima* and their cross-species applicability among waterfowl species (Anatidae). Mol. Ecol. Notes 3, 250-252.

Poulter, R., Butler, M., Ormandy, J., 1999. A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons. Gene 227, 169-179.

Raikow, R. J., 1982. Monophyly of the Passeriformes - test of a phylogenetic hypothesis. Auk 99, 431-445.

Rannala, B., Yang, Z. H., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304-311.

Sasaki, T., Takahashi, K., Nikaido, M., Miura, S., Yasukawa, Y., Okada, N., 2004. First application of the SINE (short interspersed repetitive element) method to infer phylogenetic relationships in reptiles: An example from the turtle superfamily Testudinoidea. Mol. Biol. Evol. 21, 705-715.

Schmitz, J., Ohme, M., Zischler, H., 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics 157, 777-784.

Sclater, W. L., 1930. Systema avium Aethiopicarum. Taylor and Francis, London.

Sharpe, R. B., 1883. Catalogue of the birds in the British Museum, Vol. 7. British Museum (Natural History), London.

Shedlock, A. M., Milinkovitch, M. C., Okada, N., 2000. SINE evolution, missing data, and the origin of whales. Syst. Biol. 49, 808-817.

Shedlock, A. M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. BioEssays 22, 148-160.

Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I., Okada, N., 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388, 666-670.

Sibley, C. G., Ahlquist, J. E., 1990. Phylogeny and classification of birds: a study in molecular evolution. Yale University Press, New Haven, CT.

Sibley, C. G., Monroe, B. L., 1990. Distribution and taxonomy of birds of the world. Yale University Press, New Haven, CT.

Silva, R., Burch, J. B. E., 1989. Evidence that chicken CR-1 elements represent a novel family of retroposons. Mol. Cell. Biol. 9, 3563-3566.

St. John, J., Cotter, J. P., Quinn, T. W., 2005. A recent chicken repeat 1 retrotransposition confirms the Coscoroba-Cape Barren goose clade. Mol. Phylogenet. Evol. 37, 83-90.

Stumph, W. E., Kristo, P., Tsai, M. J., O'Malley, B. W., 1981. A chicken middle-repetitive DNA-sequence which shares homology with mammalian ubiquitous repeats. Nucl. Acids Res. 9, 5383-5397.

Stumph, W. E., Baez, M., Beattie, W. G., Tsai, M. J., O'Malley, B. W., 1983. Characterization of deoxyribonucleic-acid sequences at the 5' and 3' borders of the 100-kilobase pair ovalbumin gene domain. Biochemistry 22, 306-315.

Swainson, W., 1832. Appendix, No. I; characters of genera and sub-genera hitherto undefined. *In* Swainson, W & Richardson, J. Fauna Boreali-Americana. Part 2, Birds. John Murray, London.

Swofford, D. L., 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer Associates, Sunderland, MA.

Thompson, H. S., Fotso, R., 1995. Rockfowl, the genus *Picathartes*. Bull. Afr. Bird Club 2, 25-28.

Vandergon, T. L., Reitman, M., 1994. Evolution of chicken repeat-1 (CR-1) elements - evidence for ancient subfamilies and multiple progenitors. Mol. Biol. Evol. 11, 886-898.

Watanabe, M., Nikaido, M., Tsuda, T. T., Inoko, H., Mindell, D. P., Murata, K., Okada, N., 2006. The rise and fall of the CR1 subfamily in the lineage leading to penguins. Gene 365, 57-66.

Table 1 Presence of Cor-CR-1-Loci in representative passerine taxa (+ locus-specific PCR product; * unspecific PCR product; - no PCR product). Taxonomy according to Sibley and Monroe (1990).

(Sub) Family - tribe	Species	Cor1	Cor2
Irenidae	*Irena puella* (Asian fairy bluebird)	-	-
Corvinae - Corvini	*Corvus corax* (raven)	+	+
Corvinae - Corvini	*Corvus corone* (carrion crow)	+	+
Corvinae - Corvini	*Cyanocitta stelleri* (Steller's jay)	+	+
Corvinae - Corvini	*Pica pica* (magpie)	+	+
Corvinae - Paradisaeini	*Manucodia keraudrenii* (trumpet manucode)	+	+
Corvinae - Artamini	*Gymnorhina tibicen* (white-backed (Australian) magpie)	+	+
Corvinae - Oriolini	*Oriolus chinensis* (black-naped oriole)	-	+
Picathardidae	*Picathartes gymnocephalus* (white-necked picathartes)	-	+
Picathardidae	*Picathartes oreas* (grey-necked picathartes)	-	+
Bombycillidae	*Bombycilla garrulus* (Bohemian waxwing)	-	*
Cinclidae	*Cinclus cinclus* (white-throated dipper)	-	*
Muscicapinae - Saxicolini	*Erithacus rubecula* (European robin)	-	-
Muscicapinae - Saxicolini	*Phoenicurus ochruros* (black redstart)	-	-

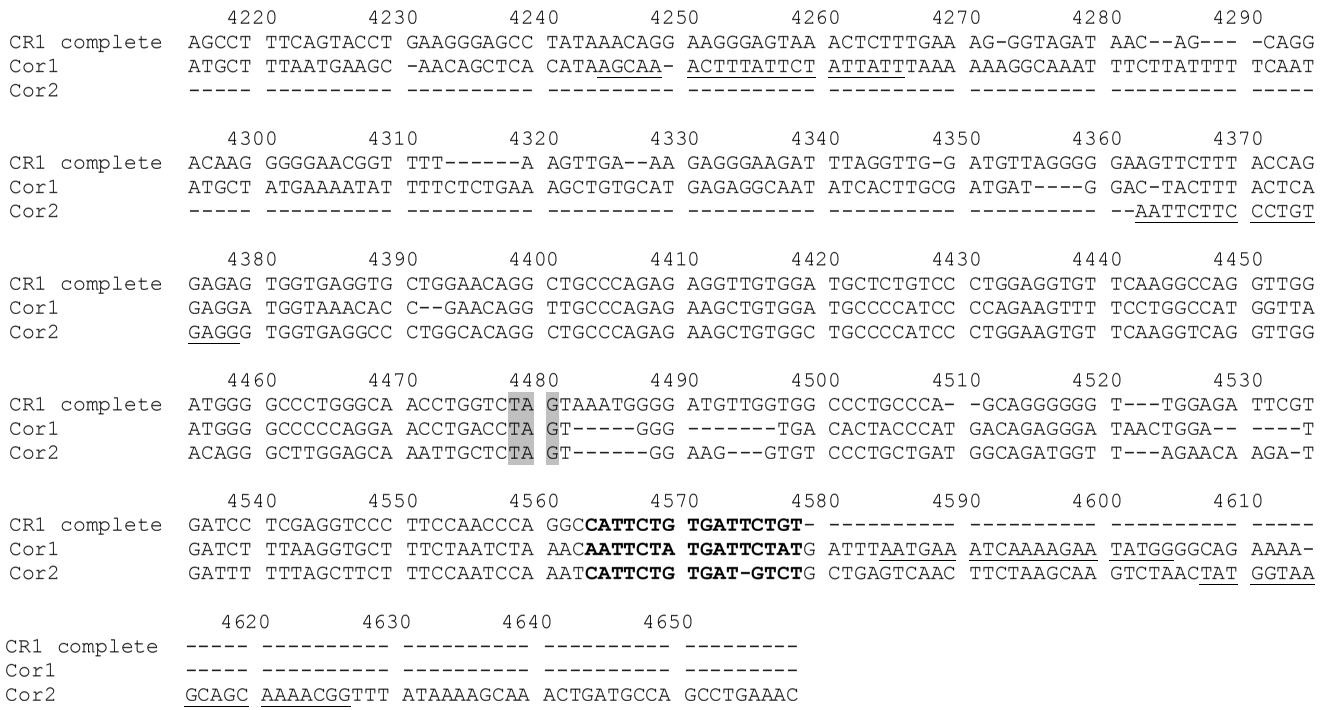


Fig. 1 CR1 clones found in the raven compared to the CR1 complete consensus sequence of the Chicken (Haas *et al*. 1997). Numbers refer to bp in this consensus sequence. End of the reverse transcriptase shaded grey, direct repeat at the 3'-end in bold and primer-locations underlined.

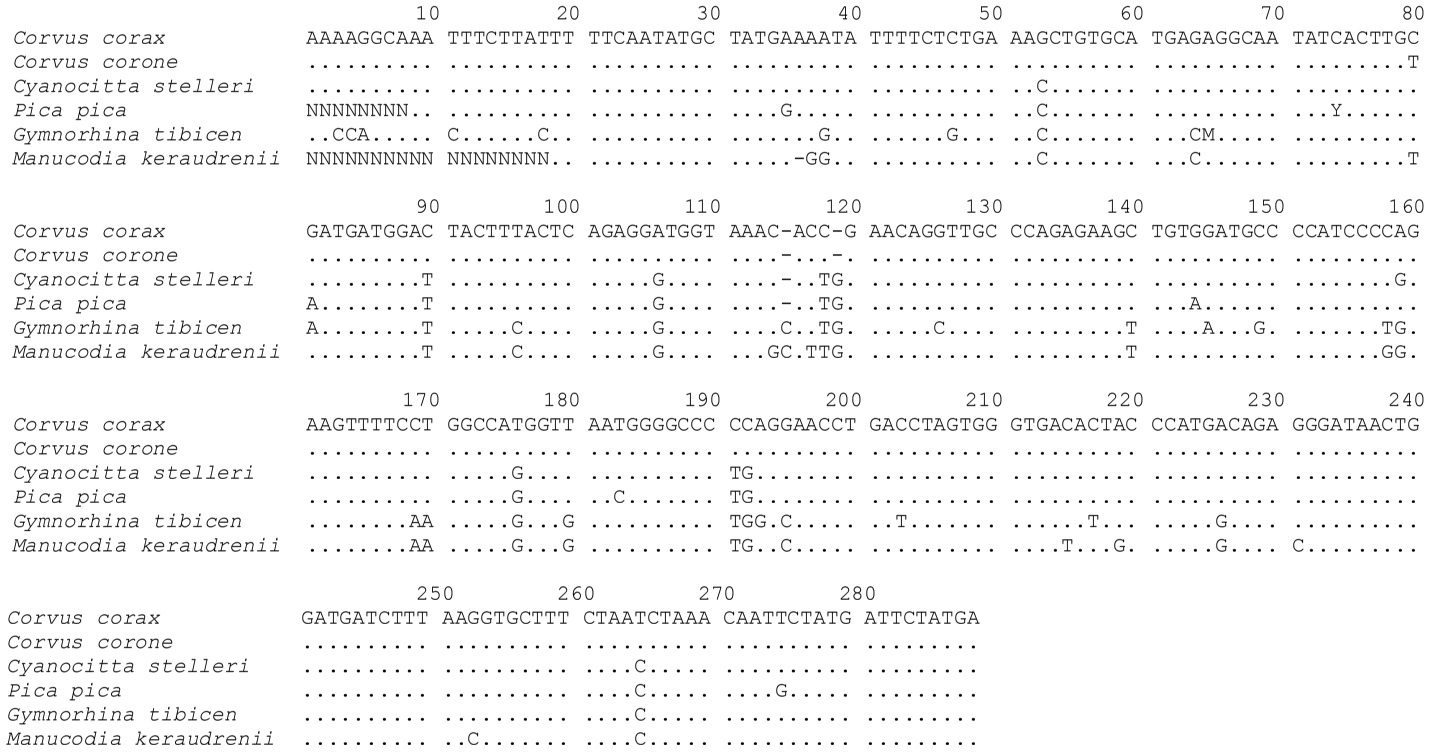


Fig. 2 Sequence alignment of the Cor1-CR1 locus.

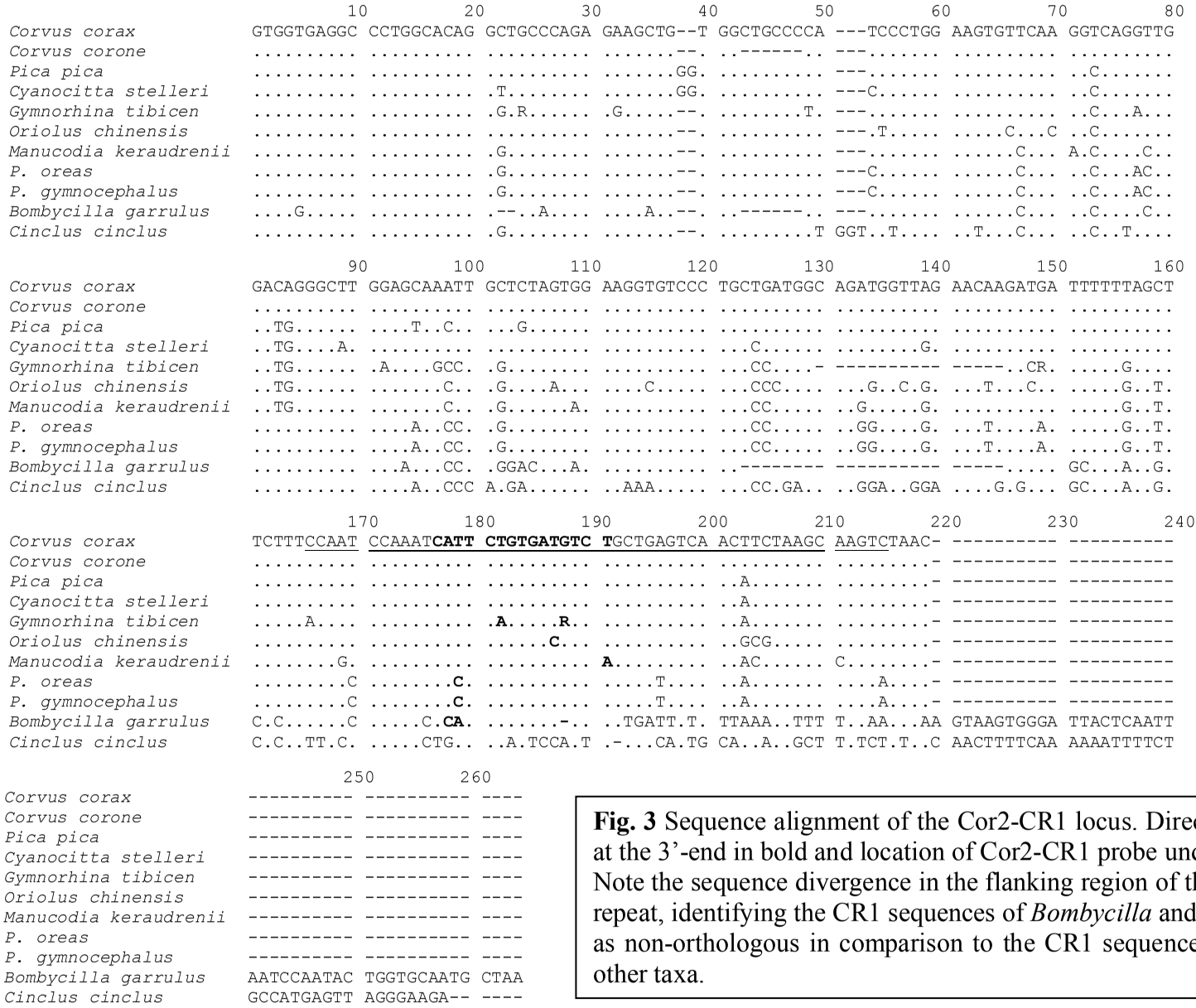


Fig. 3 Sequence alignment of the Cor2-CR1 locus. Direct repeat at the 3'-end in bold and location of Cor2-CR1 probe underlined. Note the sequence divergence in the flanking region of the direct repeat, identifying the CR1 sequences of *Bombycilla* and *Cinclus* as non-orthologous in comparison to the CR1 sequences of the other taxa.

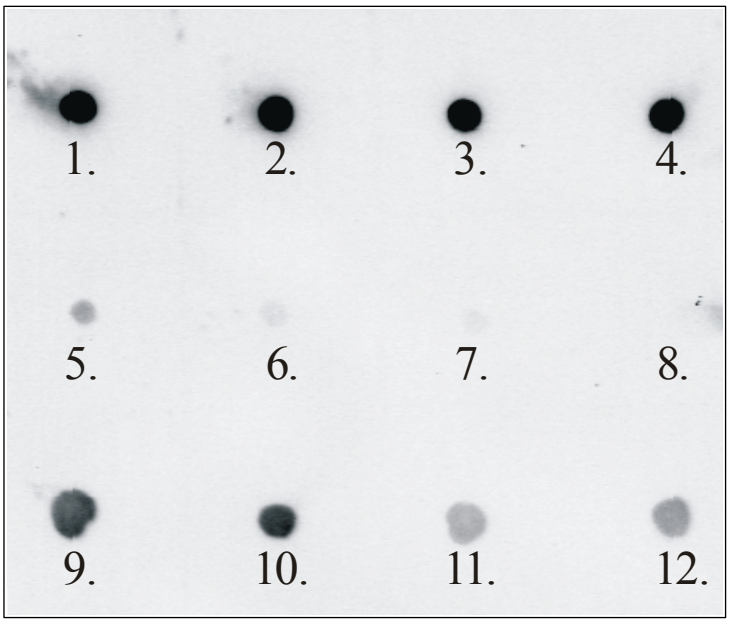


Fig. 4 Dot blot of the Cor2-CR locus. Dot 1-5 Cor2-CR1 PCR products of raven (*Corvus corax*), carrion crow (*Corvus corone*), Steller's jay (*Cyanocitta stelleri*), Bohemian waxwing (*Bombycilla garrulus*) and white-throated dipper (*Cinclus cinclus*), dot 6 and 7 Cor1-CR1 PCR products of raven and carrion crow, 8 water, 9-12 genomic DNA of carrion crow, Bohemian waxwing, black redstart (*Phoenicurus ochruros*) and great tit (*Parus major*).

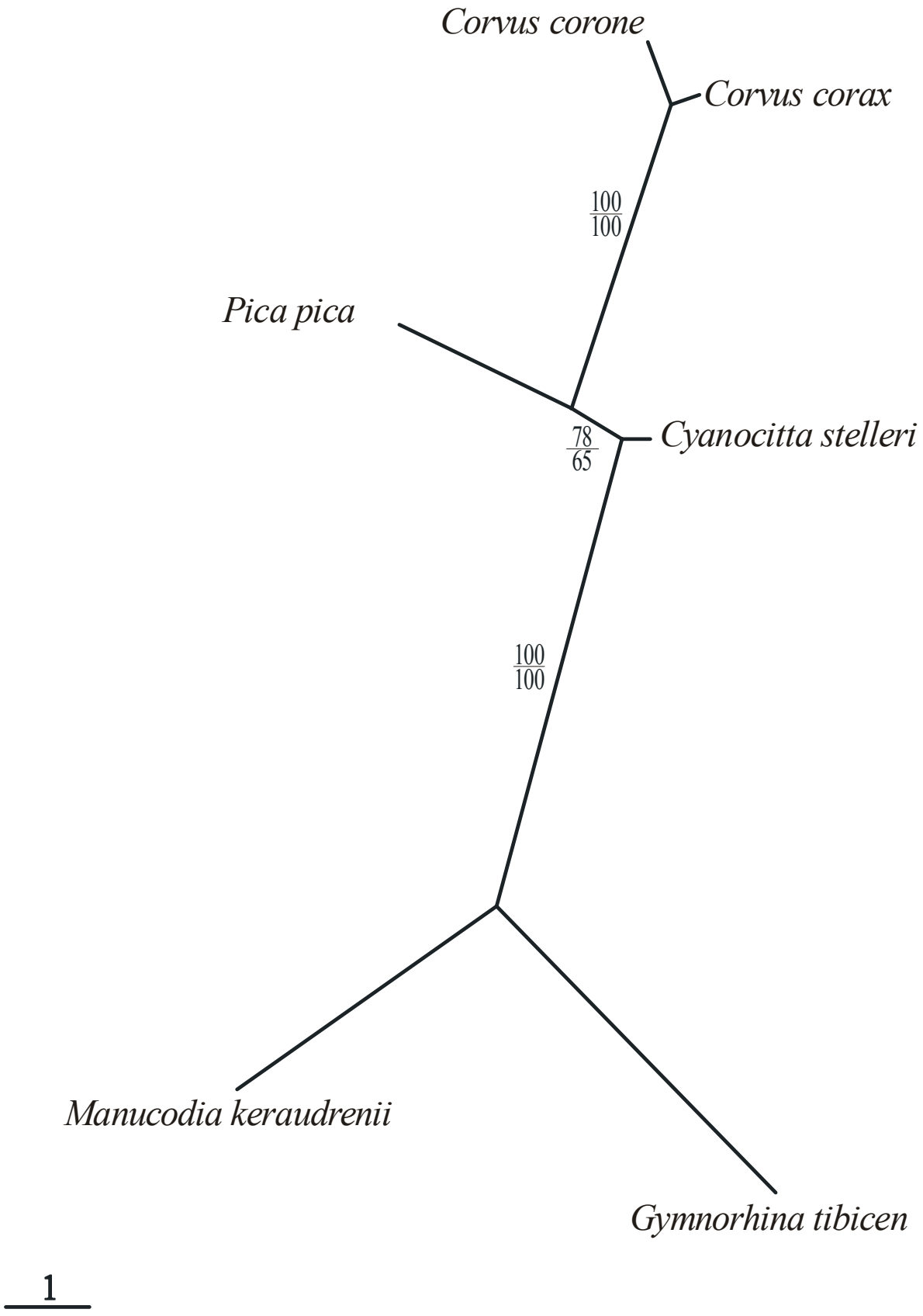


Fig. 5 Bayesian Cor1-CR1 tree with support values indicated at the branches. Estimated Bayesian posterior probabilities above and parsimony bootstrap support below the line.

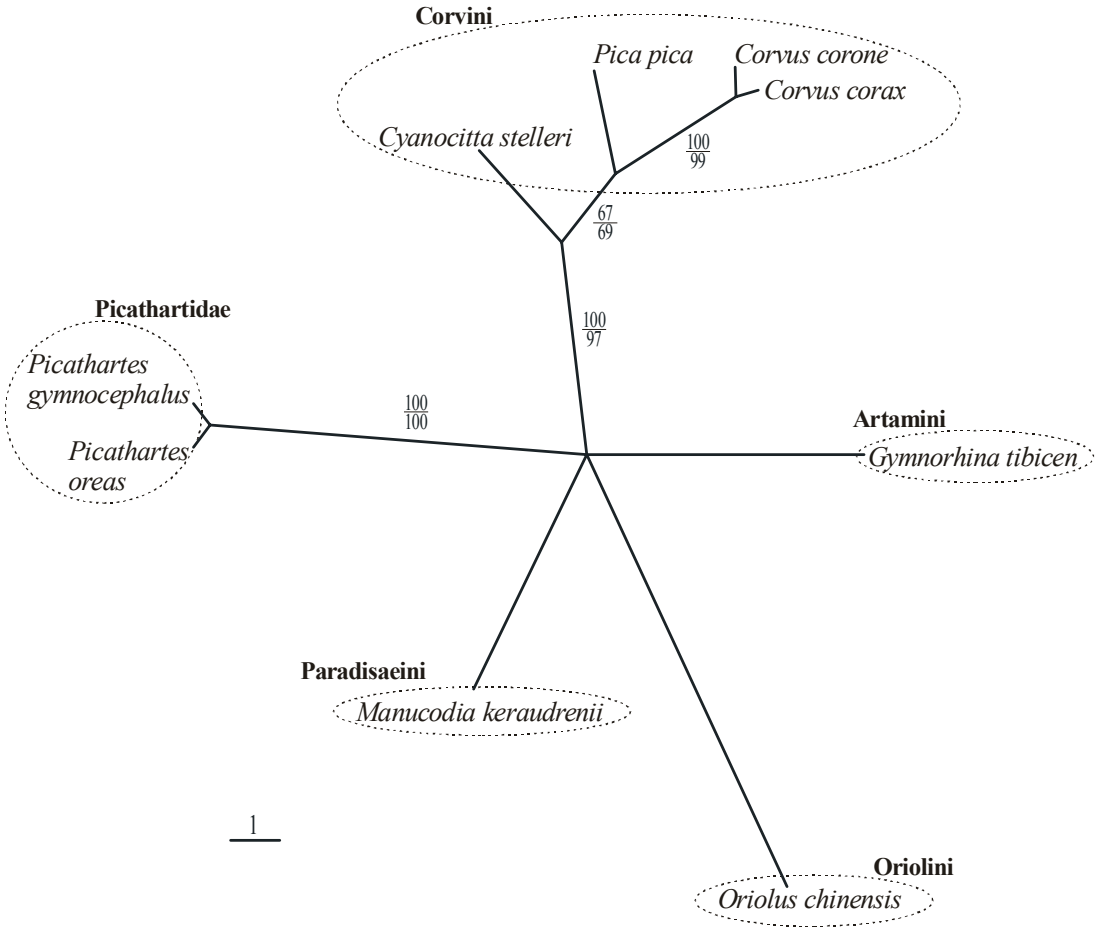


Fig. 6 Bayesian Cor2-CR1 tree with support values indicated at the branches. Estimated Bayesian posterior probabilities above and parsimony bootstrap support below the line.

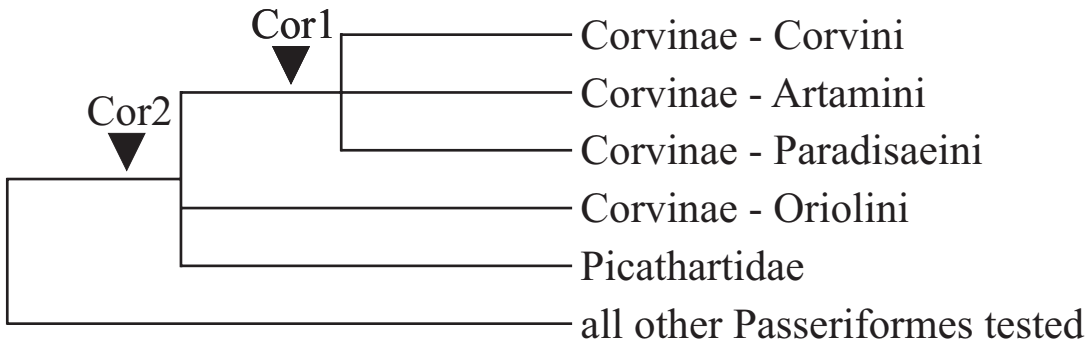


Fig. 7 Cladogram based on CR1 loci insertions found in passerine birds.

7.3 Article III:

SIMONE TREPLIN and RALPH TIEDEMANN.

Phylogenetic utility of chicken repeat 1 (CR1) retrotransposon sequences in passerine birds (Aves: Passeriformes).

Manuscript.

Abstract

The suitability of retrotransposons as apomorphic markers to infer phylogenies has repeatedly been proven. Apart from this approach, there is evidence that retrotransposon sequences themselves contain a phylogenetic signal. To investigate this specifically, we screened genomes of several species of Passeriformes for chicken repeat 1 (CR1) elements, the most widespread and important retrotransposon type in birds. We isolated seven CR1 loci and were able to amplify these loci in several species other than the source organism. Additionally, we analysed a CR1 locus found in GenBank that hitherto had been overlooked and added it to our study. Each locus was evaluated concerning sequence characteristics and the degree of saturation. A phylogenetic analysis was performed using the Bayesian approach and maximum parsimony for each locus by itself and for two combined data sets comprising species of the passerine superfamilies Muscicapoidea and Sylvioidea. We compared distances of CR1 alignments to two nuclear markers established in molecular phylogenetics for Passeriformes. We found that CR1 elements were highly variable. To investigate the phylogenetic contents of our data sets we conducted a likelihood-mapping. This study provides evidence for the phylogenetic utility of CR1 retrotransposon sequences, in addition to the classical presence/absence pattern typically scored in retrotransposon studies.

Keywords: CR1 sequences, retrotransposon, molecular phylogenetics, Passeriformes

INTRODUCTION

Whereas 40-50 % of the mammalian genome consist of interspersed repeats (IHGSC 2001; MGSC 2002), retrotransposons comprise less than 9 % of the chicken genome, the only avian genome completely analysed so far (ICGSC 2004). Retrotransposons are mobile genetic elements that are integrated into the genome via RNA intermediates. They can be divided into a viral subfamily (containing retroviruses, long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons) and a nonviral one (containing processed pseudogenes and short interspersed nuclear elements (SINEs)) (Shedlock and Okada, 2000). Chicken repeat 1 (CR1) retrotransposon insertions (Stumph *et al*., 1981) constitute the largest number of these elements (ICGSC 2004), with more than 80 % (up to 200,000 copies in the chicken genome), and is the most important non-LTR retrotransposon in birds. These elements possess an 8 bp direct repeat [typically (CATTCTRT) $(GATTCTRT)_{1-3}$ with some known variations] at the

3'-end, which can be easily detected (Silva and Burch, 1989). The first complete consensus CR1 sequence, published by Haas *et al*. (1997), contained two open reading frames (ORF), of which ORF1 presumably codes for a nucleic binding protein and ORF2 for an endonuclease and reverse transcriptase (Haas *et al*. 1997). Recently, the ICGSC (2004) described only one full-length 4.5 kb CR1 element with both intact ORFs. The first study about the evolution of CR1-elements described at least six different subfamilies (A-F) (Vandergon and Reitman, 1994). In the full genomic sequence of chicken, eleven complete CR1 source genes, subdivided into 22 subfamilies, were identified (ICGSC 2004). These results suggests an ancient origin of these elements (Vandergon and Reitman, 1994). This was confirmed, when CR1 elements were found in the genomes of other vertebrates (Chen *et al*., 1991; Fantaccione *et al*., 2004; Kajikawa *et al*., 1997; Poulter *et al*., 1999), and CR1-like elements were even reported for some invertebrate species (Albalat *et al*., 2003; Biedler and Tu, 2003; Drew and Brindley, 1997; Malik *et al*., 1999). The vast majority of elements have severely truncated 5'-ends and have lost their retrotransposition ability (Silva and Burch, 1989; Stumph *et al*., 1981).

SINE insertions have been well established as reliable apomorphic characters for phylogenetic inference (e.g., Huchon *et al*., 2002; Lum *et al*., 2000; Nikaido *et al*., 2001; Nikaido *et al*., 1999; Sasaki *et al*., 2004; Schmitz *et al*., 2001; Shedlock *et al*., 2000; Shimamura *et al*., 1997). This approach was recently applied to CR1 insertions as phylogenetic markers, i.e. one single insertion in the lactate dehydrogenase B gene was used to support the monophyly of the Coscoroba-Cape Barren goose clade within Anseriformes (St. John *et al*., 2005), and a CR1 subfamily for considering the penguin phylogeny (Watanabe *et al*., 2006). Additionally, we were able to use two CR1 elements to support a novel phylogenetic hypothesis in Passeriformes (Treplin and Tiedemann, under review). The power of retrotransposon insertions as phylogenetic markers with apparently unambiguous and homoplasy-free results is, however, compromised by the immense effort needed to find a sufficient number of phylogenetically informative retrotransposon insertions for a specific clade. In the course of these efforts, many insertions are sequenced. However, the presence of an insertion might turn out uninformative for testing a particular hypothesis, when every representative of the group under investigation might share that insertion. We argue that the sequences of these elements are potentially phylogenetically informative as well. After the insertion of a retrotransposable element at a specific locus in the genome of a common ancestor and the loss of the retrotransposable function by truncation, sequence evolution is likely not constrained by selection. Instead, these elements can be expected to evolve

according to a neutral model and, thus, can be treated as neutral molecular markers (see Treplin and Tiedemann, under review, for a first application).

To investigate the phylogenetic utility of CR1 sequences as neutral markers, we specifically searched for CR1 elements in the genomes of several species of Passeriformes. This by far largest avian taxon (comprising more than a half of all living birds Sibley and Ahlquist (1990) is well suited for such a study because the monophyly of the order itself is well established (Raikow, 1982; Sibley and Ahlquist, 1990) and phylogenetic relationships on higher taxonomic levels are strongly supported. Additionally, many studies on passerine phylogenetic relationships have been published, using large taxon samplings and several nuclear markers (e.g. Barker *et al*., 2002; Barker *et al*., 2004; Beresford *et al*., 2005; Ericson and Johansson, 2003; Irestedt *et al*., 2002; Irestedt *et al*., 2001), which facilitates the comparison of resulting phylogenetic trees of CR1 sequences to current phylogenetic hypotheses.

MATERIALS AND METHODS

We used a biotin-labelled 26 bp (5'-GMMMMGGYTKCCCRRAGARGYTGTGG-3') oligonucleotide as a probe to establish a CR1 elements-enriched DNA genomic library from a total DNA extract of different samples of representatives of Passeriformes (i.e., the great tit *Parus major*, the song thrush *Turdus philomelos*, and the European pied flycatcher *Ficedula hypoleuca*) following a standard protocol for microsatellite enrichment (Paulus and Tiedemann, 2003). This probe was designed to fit the most conserved region of CR1 elements. Recovered enriched fragments were transformed into competent *Escherichia coli* (TOPO cloning kit, Invitrogen). Recombinants were blotted onto a nylon membrane (Qiagen) and again hybridised with the CR1 probe. Positive clones were detected using the Phototope-Star chemiluminescent detection kit (New England Biolabs), sequenced with the BigDye version 3.1 Terminator Cycle Sequencing Kit (Applied Biosystems), and analysed on an AB 3100 multicapillary automatic sequencer (Applied Biosystems).

We isolated seven CR1 passerine elements (Par1 to Par3-CR1, Tur1 to Tur2-CR1, and Fic1 to Fic2-CR1) belonging to the abundant 5'-truncated type of insertions. We used these sequences to design new specific primers (Table 1). We located forward primers at the 5'-end of each clone, while reverse primers were located in the 3'-flanking region defined by the position of the CR1-specific 8 bp direct repeat identified in the clones. These primers yielded

single amplicons from their source species' genomic DNA. They were also used for PCR-amplifications on genomic DNA from different passerine birds that are supposed to be closely related (following Sibley and Monroe (1990) (Table2).

Additionally to our cloned CR1 elements, we identified a hitherto undescribed CR1 insertion in GenBank that is associated with a nuclear pseudogene of the mitochondrial cytochrome *b* (numt) in Darwin's finches. We suspect that this insertion was not recognised as CR1 by the original authors (Sato *et al*., 2001), because this insertion is in reverse complement direction compared to the numt. We named it 'Darfin-CR1', according to the method of naming a CR1 element after the species in which it was found. As the primers of Sato *et al*. (2001) failed to amplify this insertion in species other than Darwin's finches, we developed new primers (Table 1).

PCR-amplifications were performed in a total volume of 37.5 µl, containing 1 mM Tris-HCl, pH 9.0, 5 mM KCl, 0.15 mM $MgCl_2$, 0.05 mM of each dNTP, 0.13 µM of both forward and reverse primers and 0.75 U *Taq* polymerase (Qbiogene) in a Biometra T3000 thermocycler according to the following reaction profile: 1 cycle at 96°C for 5 min, 40 cycles at 96°C for 1 min 30 s, locus-specific annealing temperature (see Table 1) for 1 min 15 s, 72°C for 1 min 30 s, and a final extension at 72°C for 10 min. Cycle sequencing reactions were performed using the forward and reverse primers.

Sequences were aligned in the BioEdit Sequence Alignment Editor (Hall, 1999) and analysed phylogenetically both as single locus data sets and combined analyses (named Mus-CR1 and Syl-CR1 for Muscicapoidea and Sylvioidea, respectively) where several CR1 loci were found in the same species (Table 2). For the combined analyses, nucleotides were treated as unknown where locus specific PCR failed to amplify the CR1 element. A chi-square test of homogeneity of base frequencies across taxa was used for each gene to test for variation in the base frequencies between the OTUs. Data sets of the different genes were tested for heterogeneity using the partition homogeneity test (Farris *et al*., 1995), implemented in PAUP* (Swofford, 2001), to assess the appropriateness of combining the data partitions. We conducted a test between each pair of gene partitions using 1,000 replicates for each test. Maximum parsimony analyses (MP, Farris *et al*., 1970) were performed using the heuristic search option with the TBR-branch-swapping algorithm in PAUP* 4.0b10 (Swofford, 2001). Robustness of the phylogenetic hypotheses were evaluated by bootstrapping (Felsenstein, 1985) with 1,000 and 10,000 replicates. We used MrModeltest version 2.2 (Nylander, 2004) to identify the best model of sequence evolution for both datasets. Data were analysed under a

maximum likelihood criterion (ML, Felsenstein, 1981) in a Bayesian framework (Huelsenbeck *et al*., 2000; Larget and Simon, 1999; Mau and Newton, 1997; Mau *et al*., 1999; Rannala and Yang, 1996) using MrBayes 3.0b4 (Huelsenbeck and Ronquist, 2001). All priors were set according to the chosen model and according to each partition (Table 3). Four Markov chains, three heated and one cold, were started from a random tree and all four chains ran simultaneously for 1,000,000 generations (for the combined analysis and for each separate CR1 respectively), with trees being sampled every 100 generations for a total of 10,001 trees. After the likelihood of the trees of each chain converged, the first trees were discarded as *burn in* (150 for Mus-CR1 and 200 for Syl-CR1). The majority-rule consensus tree containing the posterior probabilities of the phylogeny was determined afterwards. In the combined data set analyses, the common starling *Sturnus vulgaris* was chosen as an outgroup.

To test for saturation, pairwise transition and transversion sequence distances were plotted against total sequence distances. The variability of CR1 elements was evaluated by comparison of pairwise p-distances of CR1 sequences to sequences of the two nuclear marker genes proven useful in resolving passerine phylogenies, ZENK (Chubb, 2004; Treplin *et al.*, submitted), and RAG-1 (e.g. Barker *et al.*, 2002; Barker *et al.*, 2004; Beresford *et al.*, 2005; Ericson and Johansson, 2003; Irestedt *et al.*, 2002; Irestedt *et al.*, 2001). To visualise the content of phylogenetic information of the Mus-CR1 and Syl-CR1 alignments, we used the likelihood-mapping method (Strimmer and von Haeseler, 1997), which is based on the analysis of quartet puzzling, included in TREE-PUZZLE 5.0, (Schmidt *et al*., 2002) with the model set to HKY (Hasegawa *et al*., 1985).

RESULTS

In the CR1 enriched genomic libraries we were able to identify seven CR1 elements by direct comparisons with the complete CR1 consensus sequence described by Haas *et al*. (1997). Sequences of the flanking regions were excluded from all analyses. All sequences are deposited in GenBank (accession nos. **XXXXXX-XXXXXX**, cf. Table 2), and final alignment lengths are shown in Table 3. Translation into amino acids showed sequence homology to reverse transcriptase, but – as expected for truncated CR1 elements – genes were not functional, as they frequently contained indels shifting the reading frame. All CR1 sequences of the Blyth's reed-warbler (*Acrocephalus dumetorum*) and the icterine warbler (*Hippolais icterina*) were identical. Amplification of the Darfin-CR1 yielded products in

representatives of all major passerine taxa, other than the Muscicapoidea and Sylvioidea, and all sequences were included in distance calculations, instead of adding just 21 and 10 the Syl-CR1 and the Mus-CR1 data set. Sequences of the Darfin-CR1 element were added to both Mus-CR1 and Syl-CR1 combined data sets. Overall, uncorrected pairwise divergence among passerine birds varied from 0.0 within genera to 0.364 in the Darfin-CR1 data set (Table 3). The chi-square test of homogeneity of base frequencies across taxa did not show significant heterogeneity (p=0.99 to 1.0) in any data set. The partition homogeneity test showed no significant heterogeneity among all data partitions in the combined data sets Mus-CR1 and Syl-CR1 of all loci (Table 4). Despite the high variability with most loci having a proportion of invariable sites of zero, the number of parsimonious informative sites, ranging from 7 to 28 % of the complete sequence length, is comparably low. Additionally, plotting transition and transversion sequence distances over total distances did not detect any saturation (Fig. 1). Homoplasy seems to have low impact as indicated by the high values of the homoplasy indices (Table 3). The high variability is illustrated by comparison of the CR1 loci with two nuclear markers, ZENK and RAG-1 (Fig. 2). Correlation of distances compared to ZENK was significant for every locus, whereas two loci (Par3-CR1 and Fic1-CR1) showed no significant correlation of distances to RAG-1. In all but two significantly correlated loci (Par1 vs. RAG-1, and Par2 vs. ZENK), the CR1-variability was higher compared with the nuclear genes. Figure 3 shows the results of the likelihood mapping of the Syl-CR1 (A) and Mus-CR1 (B) data sets. This method defines seven areas of attraction (the lower triangles in Fig. 3): The quartet-puzzling method yields three possible fully resolved tree topologies by comparing four sequences, represented by the corners of the triangle. A star phylogeny or sequences that are too short sometimes make it impossible to resolve the phylogenetic relationships of four sequences. Thus, the mid triangles represent the region where the star-like tree is the optimal tree. The rectangles between the corners of the triangle represent the situation when two fully resolved trees could not be distinguished. The number of fully resolved trees is much higher in the Mus-CR1 (87.7 %) than in the Syl-CR1 data set (70.0 %). With 21.4 %, the Syl-CR1 data set shows a relatively high likelihood of star-like evolution, whereas the likelihood of the Mus-CR1 data set is much lower with 6.7 % of all quartet points. The percentage of quartets mapped into the regions, where two topologies could not be distinguished, is comparable in both cases (8.6 % and 5.7 % for Syl- and Mus-CR1, respectively).

Single locus phylogenetic analyses of our CR1 loci suffered from sequences being too short and having low phylogenetic informative contents and yielded no consistent

phylogenetic trees (data not shown). Analyses of the two combined CR1 data sets resulted in phylogenetic trees that differed in their resolution and nodal support.

In the Syl-CR1 MP analysis, only closely related taxa were resolved (Fig. 4), whereas the Bayesian phylogenetic tree resolved more nodes but lacked significant support, apart from closely related taxa. Whenever more than one representative of a genus was included, they were resolved as a strongly supported monophylum with both types of analysis. The Bohemian waxwing (*Bombycilla garrulus*) and the goldcrest (*Regulus regulus*) form a monophyletic clade. The tits (*Parus* spp.) are separated from the remaining Sylvioidea with the winterwren (*Troglodytes troglodytes*) as their closest relative. Phylogenetic relationships within Sylvioidea (sensu Alström *et al.*, 2006) are poorly resolved: the Eurasian skylark (*Alauda arvensis*) is supported as the most basal branch, followed by two sister clades consisting of (1) bulbuls (*Pycnonotus* sp.), the Blyth's reed-warbler and the icterine warbler, and the northern house-martin (*Delichon urbica*), and (2) the willow warbler (*Phylloscopus trochilus*), the common grasshopper-warbler (*Locustella naevia*), the short-toed tree-creeper (*Certhia brachydactyla*), and a trichotomy consisting of the African yellow white-eye (*Zosterops senegalensis*), the white-necked laughingthrush (*Garrulax strepitans*), and the greater whitethroat (*Sylvia communis*).

In comparison, both trees based on the Mus-CR1 data set are well resolved and congruent (Fig. 5). The subfamilies Turdinae and Muscicapinae, as well as the muscicapine tribes Muscicapini and Saxicolini could be distinguished as monophyletic clades. The two redstarts (*Phoenicurus* spp.) form their own clade within the Saxicolini. The European pied flycatcher (*Ficedula hypoleuca*) and the European robin (*Erithacus rubecula*) are sister taxa and together closer related to the whinchat (*Saxicola rubetra*) than to the thrush nightingale (*Luscinia luscinia*).

DISCUSSION

Characteristics of CR1 sequences and conditions for a use as phylogenetic markers

Although many studies have been published about CR1 retrotransposons since they were discovered in 1981 (Stumph *et al.*, 1981), sequence characteristics and evolution of the same element after an insertion event at a specific locus have– to our knowledge – not been subject of investigation. Although less is known about the mechanism of retrotransposition of

non-LTR retrotransposons (Kazazian, 2004) than of LTR retrotransposons (reviewed by Wilhelm and Wilhelm, 2001), evolution of CR1 elements was supposed to trace back to at least eleven 'source genes' (ICGSC 2004). The relative abundance of CR1 elements and the ongoing increase in estimation of the number of copies in the chicken genome (i.e. from 30,000 (Burch *et al*., 1993), over 100,000 (Vandergon and Reitman, 1994) to 200,000 (ICGSC 2004)) provide the possibility of utilising them as phylogenetic markers. It was recently demonstrated that CR1 sequences contain phylogenetic signals (Treplin and Tiedemann, under review). The significance of the contained information is depending on the time span from the insertion event to the recent observed character state. Our single locus sequences showed a high variability among taxa. The single locus data sets, however, failed to resolve phylogenetic relationships unambiguously due to low numbers of informative characters and comparably short sequence lengths. The highly conserved regions of CR1 elements throughout different CR1 families and distantly related taxa pointed to an evolution under functional constraints, i.e. these conserved regions may act as transcriptional silencers and protein binding sites (Chen *et al*., 1991). Apart from these conserved regions, the other parts diverged considerably (Chen *et al*., 1991; Haas *et al*., 2001). This suggests an ancient origin of these elements, as sequences of the progenitor and the newly inserted copies should be nearly identical (Deininger *et al*., 1992; Kido *et al*., 1995). During incomplete reverse transcription from the 3' end, the functional constraints are relaxed when CR1 elements lose their retrotransposition ability. Consequently, neutral point substitutions could accumulate randomly in the genome in the large number of truncated CR1 elements, varying in length (Kido *et al*., 1995; Webster *et al*., 2006). The loci described here have obviously lost their function and developed a high variability among closely related taxa due to neutral evolution, e.g. 3.2 times higher in Fic2-CR1, compared to ZENK (Fig. 2c).

Generally, the substitution rate of CR1 sequences is correlated to that of the sequences of the two nuclear genes (Fig. 2). Despite the high variability, the problem of noise in fast and constantly evolving neutral markers does not play an important role, as no saturation was detected in the CR1 loci (Fig. 1), and the transition/transversion ratios did not indicate a high rate of multiple substitutions (Table 3). Saturation has been found to occur in molecular markers with higher variability, e.g. the mitochondrial cytochrome *b* gene, compromising its suitability for higher-level systematics in Passeriformes (e.g. Chikuni *et al.*, 1996; Edwards *et al.*, 1991; Edwards and Wilson, 1990). Additionally, all three homoplasy indices are on the same level or even higher than those of the nuclear gene markers used in Treplin *et al*. are (ZENK, RAG-1, RAG-2, c-*myc*; submitted) (Table 3).

The results of the likelihood-mapping indicate a higher content of phylogenetic signal in the Mus-CR1 data set than in the Syl-CR1 data set. Although likelihood-mapping results are influenced by sequence lengths (Strimmer and von Haeseler, 1997), this does not apply here, as both combined data sets comprise the same length. The total amount of 87.7 % fully resolved quartets in the Mus-CR1 data set could lead to a highly resolved phylogenetic tree. Despite the possibility of likelihood-mapping producing false positives (Nieselt-Struwe and von Haeseler, 2001), the tree based on the Mus-CR1 data set (Fig. 5) follows this expectation, as well as the tree based on the Syl-CR1 data set (Fig. 4). Altogether, the aligned sequence data display informative phylogenetic signal. Probably, additional sequences of CR1 loci would increase the phylogenetic signal.

Application as phylogenetic markers

Phylogenetic relationships in the Bayesian tree of the Syl-CR1 data set are generally in good agreement with recent phylogenetic hypotheses. Excluding the Paridae from the Sylvioidea (*sensu* Alström *et al.*, 2006) was strongly supported in a recent study (Treplin *et al*. submitted) and again is corroborated by this study. The position of the Eurasian skylark *Alauda arvensis* at the base of the Sylvioidea (*sensu* Alström *et al.*, 2006) confirms this phylogenetic position which had already been proposed by several other studies (Alström *et al.*, 2006; Barker *et al.*, 2004; Beresford *et al.*, 2005; Ericson and Johansson, 2003; Fuchs *et al.*, 2006). The Syl-CR1 tree also contains some ambiguous clades, such as the distribution of the wrens, tree-creepers and nuthatches, (Certhiidae, *Troglodytes* sp. and *Certhia* sp., and Sittidae, *Sitta* sp.). These taxa, placed among unlikely clades without strong support, are believed to form a monophyletic clade and to be more closely related to the Muscicapoidea (Barker *et al*., 2004; Beresford *et al*., 2005; Cibois and Cracraft, 2004). Our CR1 sequences remarkably failed to define such a clade. Phylogeny within the Sylvioidea (*sensu* Alström *et al*.) has been difficult to elucidate (Alström *et al.*, 2006; Jønsson and Fjeldså, 2006) and this taxon has appeared to be the least resolved group in the recently published supertree of Passerida in a metaanalysis based on 99 studies (Jønsson and Fjeldså, 2006). The present analysis again cannot fully resolve this clade. Polyphyly of the Acrocephalinae (*sensu* Sibley and Ahlquist, 1990) and exclusion from the Sylviidae (*sensu* Sibley and Ahlquist, 1990) has been suggested by several authors (e.g. Alström *et al.*, 2006; Sefc *et al.*, 2003) and is also confirmed in this study by the separation of both the willow warbler (*Phylloscopus trochilus*) and the common grashopper-warbler from the Blyth's reed-warbler/icterine warbler clade.

The established groups within the Muscicapidae (*sensu* Sibley and Ahlquist, 1990) are well resolved and congruent in both analyses (Fig. 4). The blackbird *Turdus merula* and the hermit thrush *Catharus guttatus* form a strongly supported monophyletic clade, which represents the Turdinae, defined by Sibley and Ahlquist, (1990) as sister to the Muscicapinae. This clade is in congruence to many other sequence-based studies (e.g. Barker *et al.*, 2004; Beresford *et al.*, 2005; Cibois and Cracraft, 2004; Jønsson and Fjeldså, 2006). Division of the Muscicapinae into the two tribes Muscicapini and Saxicolini is confirmed by our data set regarding the only representative of Muscicapini, the spotted flycatcher *Muscicapa striata*. The European pied flycatcher was classified as a Muscicapini by Sibley and Monroe (1990), but it has recently been shown that instead it belongs to the Saxicolini (e.g. Treplin *et al*., submitted). Phylogenetic relationships within the Saxicolini are difficult to evaluate, because hardly any comprehensive study exists dealing with that clade. Higher-level studies typically include only few saxicoline taxa. The supertree analyses of Jønsson and Fjeldså (2006) combined different studies and hence yielded a higher taxon density. In this supertree, it is apparent that the two species European pied flycatcher and European robin are basal representatives of the Saxicolini. This had also been found by Cibois and Cracraft (2004), who additionally showed a sister taxon relationship of the genera *Phoenicurus* and *Luscinia*. Our analyses therefore agree with recently hypothesised higher-level phylogenetic relationships.

The two data sets of combined CR1 sequences, though analysing phylogenetic relationships on comparably same taxonomical levels, showed remarkable differences with regard to resolution, congruence, and support. These differences may stem from unequal rates of evolution in these two groups, indicated by the different phylogenetic informative signal contained in these markers (see above). Sheldon and Gill (1996) summarised the long history of difficulties investigating sylvioid relationships. The Sylvioidea comprise the second largest group of oscine birds (*sensu* Sibley and Monroe, 1990). Radiation and speciation might have been particularly rapid in the Sylvioidea, leading to a large group of closely related species (compared to the obviously less diversified group of Muscicapoidea), as indicated by the very short branch lengths in recent phylogenetic analyses (Treplin *et al*., submitted) and the weak resolution in the metaanalysis of Jønsson and Fjeldså (2006). With short intervals between branching events, evolving clades likely acquired few – if any – synapomorphies (Lanyon, 1988).

Conclusion

This study proves the suitability of CR1 loci as phylogenetic markers. The benefit of available sequences, possibly unsuitable for classical retrotransposon studies with synapomorphic character approaches, is worth utilising as standard sequence based analyses. As there is such a high number of elements in the birds' genome, generating a larger data set (i.e. more loci/longer sequences and further taxa included) than those included in the present study, could definitely contribute to the ongoing debate in passerine phylogenies. In particular, CR1 sequences, as a source of multilocus nuclear phylogenetic markers, are potentially less prone to possible lineage sorting effects, which can cause gene tree/species tree incongruencies in studies based on single or few loci.

ACKNOWLEDGEMENTS

The authors would like to thank the following institutions and people for providing needed samples: University of Kiel, Haustierkunde (H. Luttmann and J. Oesert), Zoological Garden Berlin (R. Reinhard), Tierpark Berlin (M. Kaiser), Dieter Wallschläger, and Malte Mews. We thank Christoph Bleidorn and Susanne Hauswaldt for valuable comments on the manuscript.

REFERENCES

Albalat, R., Permanyer, J., Martinez-Mir, C., Gonzalez-Angulo, O., Gonzalez-Duarte, R., 2003. The first non-LTR retrotransposon characterised in the cephalochordate amphioxus, BfCR1, shows similarities to CR1-like elements. Cell. Mol. Life Sci. 60, 803-809.

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.

Barker, F. K., Barrowclough, G. F., Groth, J. G., 2002. A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. Proc. R. Soc. Lond. B Biol. Sci. 269, 295-308.

Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., Cracraft, J., 2004. Phylogeny and diversification of the largest avian radiation. Proc. Natl. Acad. Sci. U.S.A. 101, 11040-11045.

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 Biol. Sci. 272, 849-858.

Biedler, J., Tu, Z. J., 2003. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. Mol. Biol. Evol. 20, 1811-1825.

Burch, J. B. E., Davis, D. L., Haas, N. B., 1993. Chicken repeat-1 elements contain a *pol*-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. Proc. Natl. Acad. Sci. U.S.A. 90, 8199-8203.

Chen, Z. Q., Ritzel, R. G., Lin, C. C., Hodgetts, R. B., 1991. Sequence conservation in avian CR-1 - an interspersed repetitive DNA family evolving under functional constraints. Proc. Natl. Acad. Sci. U.S.A. 88, 5814-5818.

Chikuni, K., Minaka, N., Ikenaga, H., 1996. Molecular phylogeny of some Passeriformes, based on cytochrome *b* sequences. J. Yamashina Inst. Ornithol. 28, 1-8.

Chubb, A. L., 2004. Nuclear corroboration of DNA-DNA hybridization in deep phylogenies of hummingbirds, swifts, and passerines: the phylogenetic utility of ZENK (ii). Mol. Phylogenet. Evol. 30, 128-139.

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.

Deininger, P. L., Batzer, M. A., Hutchison, C. A., Edgell, M. H., 1992. Master genes in mammalian repetitive DNA amplification. Trends Genet. 8, 307-311.

Drew, A. C., Brindley, P. J., 1997. A retrotransposon of the non-long terminal repeat class from the human blood fluke *Schistosoma mansoni*. Similarities to the chicken-repeat-1-like elements of vertebrates. Mol. Biol. Evol. 14, 602-610.

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.

Edwards, S. V., Arctander, P., Wilson, A. C., 1991. Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. Proc. R. Soc. Lond. B Biol. Sci. 243, 99-107.

Ericson, P. G. P., Johansson, U. S., 2003. Phylogeny of Passerida (Aves: Passeriformes) based on nuclear and mitochondrial sequence data. Mol. Phylogenet. Evol. 29, 126-138.

Fantaccione, S., Russo, C., Palomba, P., Rienzo, M., Pontecorvo, G., 2004. A new pair of CR1-like LINE and tRNA-derived SINE elements in *Podarcis sicula* genome. Gene 339, 189-198.

Farris, J. S., Kluge, A. G., Eckardt, M. J., 1970. A numerical approach to phylogenetic systematics. Syst. Zool. 19, 172-189.

Farris, J. S., Kallersjo, M., Kluge, A. G., Bult, C., 1995. Constructing a significance test for incongruence. Syst. Biol. 44, 570-572.

Felsenstein, J., 1981. Evolutionary trees from DNA-sequences - a maximum-likelihood approach. J. Mol. Evol. 17, 368-376.

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

Fuchs, J., Fjeldså, J., Bowie, R. C., 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.

Haas, N. B., Grabowski, J. M., Sivitz, A. B., Burch, J. B. E., 1997. Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames. Gene 197, 305-309.

Haas, N. B., Grabowski, J. M., North, J., Moran, J. V., Kazazian, H. H., Burch, J. B. E., 2001. Subfamilies of CR1 non-LTR retrotransposons have different 5 ' UTR sequences but are otherwise conserved. Gene 265, 175-183.

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

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

Huchon, D., Nikaido, M., Okada, N., 2002. Use of SINE retroposon in phylogeny. Médecine Sciences 18, 1276-1281.

Huelsenbeck, J. P., Rannala, B., Masly, J. P., 2000. Accommodating phylogenetic uncertainty in evolutionary studies. Science 288, 2349-2350.

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

ICGSC - International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695-716.

IHGSC - International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Irestedt, M., Johansson, U. S., Parsons, T. J., Ericson, P. G. P., 2001. Phylogeny of major lineages of suboscines (Passeriformes) analysed by nuclear DNA sequence data. J. Avian Biol. 32, 15-25.

Irestedt, M., Fjeldså, J., Johansson, U. S., Ericson, P. G. P., 2002. Systematic relationships and biogeography of the tracheophone suboscines (Aves: Passeriformes). Mol. Phylogenet. Evol. 23, 499-512.

Jønsson, K. A., Fjeldså, J., 2006. A phylogenetic supertree of oscine passerine birds (Aves: Passeri). Zool. Scr. 35, 149-186.

Kajikawa, M., Ohshima, K., Okada, N., 1997. Determination of the entire sequence of turtle CR1: The first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. Mol. Biol. Evol. 14, 1206-1217.

Kazazian, H. H., 2004. Mobile elements: drivers of genome evolution. Science 303, 1626-1632.

Kido, Y., Saitoh, M., Murata, S., Okada, N., 1995. Evolution of the active sequences of the HpaI short interspersed elements. J. Mol. Evol. 41, 986-995.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. J. Mol. Evol. 16, 111-120.

Lanyon, S. M., 1988. The stochastic mode of molecular evolution - what consequences for systematic investigations. Auk 105, 565-573.

Larget, B., Simon, D. L., 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16, 750-759.

Lum, J. K., Nikaido, M., Shimamura, M., Shimodaira, H., Shedlock, A. M., Okada, N., Hasegawa, M., 2000. Consistency of SINE insertion topology and flanking sequence tree: quantifying relationships among cetartiodactyls. Mol. Biol. Evol. 17, 1417-1424.

Malik, H. S., Burke, W. D., Eickbush, T. H., 1999. The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. 16, 793-805.

Mau, B., Newton, M. A., 1997. Phylogenetic inference for binary data on dendograms using Markov chain Monte Carlo. J. Comp. Graph. Stat. 6, 122-131.

Mau, B., Newton, M. A., Larget, B., 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Biometrics 55, 1-12.

MGSC - Mouse Genome Sequencing Consortium, 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520-562.

Nieselt-Struwe, K., von Haeseler, A., 2001. Quartet-mapping, a generalization of the likelihood-mapping procedure. Mol. Biol. Evol. 18, 1204-1219.

Nikaido, M., Rooney, A. P., Okada, N., 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. U.S.A. 96, 10261-10266.

Nikaido, M., Matsuno, F., Hamilton, H., Brownell, R. L., Cao, Y., Ding, W., Zuoyan, Z., Shedlock, A. M., Fordyce, R. E., Hasegawa, M., Okada, N., 2001. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. Proc. Natl. Acad. Sci. U.S.A. 98, 7384-7389.

Nylander, J. A. A., 2004. MrModeltest v2. Program distributed by the author.

Page, R. D. M., Holmes, E. C., 1998. Molecular evolution. A phylogenetic approach. Blackwell Science, Oxford.

Paulus, K. B., Tiedemann, R., 2003. Ten polymorphic autosomal microsatellite loci for the Eider duck *Somateria mollissima* and their cross-species applicability among waterfowl species (Anatidae). Mol. Ecol. Notes 3, 250-252.

Poulter, R., Butler, M., Ormandy, J., 1999. A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons. Gene 227, 169-179.

Raikow, R. J., 1982. Monophyly of the Passeriformes - test of a phylogenetic hypothesis. Auk 99, 431-445.

Rannala, B., Yang, Z. H., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304-311.

Sasaki, T., Takahashi, K., Nikaido, M., Miura, S., Yasukawa, Y., Okada, N., 2004. First application of the SINE (short interspersed repetitive element) method to infer phylogenetic relationships in reptiles: an example from the turtle superfamily Testudinoidea. Mol. Biol. Evol. 21, 705-715.

Sato, A., Tichy, H., O'hUigin, C., Grant, P. R., Grant, B. R., Klein, J., 2001. On the origin of Darwin's finches. Mol. Biol. Evol. 18, 299-311.

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

Schmitz, J., Ohme, M., Zischler, H., 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics 157, 777-784.

Sefc, K. M., Payne, R. B., Sorenson, M. D., 2003. Phylogenetic relationships of African sunbird-like warblers: moho (*Hypergerus atriceps*), green hylia (*Hylia prasina*) and tit-hylia (*Pholidornis rushiae*). Ostrich 74, 8-17.

Shedlock, A. M., Milinkovitch, M. C., Okada, N., 2000. SINE evolution, missing data, and the origin of whales. Syst. Biol. 49, 808-817.

Shedlock, A. M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. BioEssays 22, 148-160.

Sheldon, F. H., Gill, F. B., 1996. A reconsideration of songbird phylogeny, with emphasis on the evolution of titmice and their sylvioid relatives. Syst. Biol. 45, 473-495.

Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I., Okada, N., 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388, 666-670.

Sibley, C. G., Ahlquist, J. E., 1990. Phylogeny and classification of birds: a study in molecular evolution. Yale University Press, New Haven, CT.

Sibley, C. G., Monroe, B. L., 1990. Distribution and taxonomy of birds of the world. Yale University Press, New Haven, CT.

Silva, R., Burch, J. B. E., 1989. Evidence that chicken CR-1 elements represent a novel family of retroposons. Mol. Cell. Biol. 9, 3563-3566.

St. John, J., Cotter, J. P., Quinn, T. W., 2005. A recent chicken repeat 1 retrotransposition confirms the Coscoroba-Cape Barren goose clade. Mol. Phylogenet. Evol. 37, 83-90.

Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. U.S.A. 94, 6815-6819.

Stumph, W. E., Kristo, P., Tsai, M. J., O'Malley, B. W., 1981. A chicken middle-repetitive DNA-sequence which shares homology with mammalian ubiquitous repeats. Nucl. Acids Res. 9, 5383-5397.

Swofford, D. L., 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer Associates, Sunderland, MA.

Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., Tiedemann, R., Looking for the 'best' marker: songbird (Aves: Passeriformes) phylogeny based on sequence analyses of several unlinked nuclear loci, submitted. Syst. Biol.

Treplin, S., Tiedemann, R., Specific chicken repeat 1 (CR1) retrotransposon insertion suggests phylogenetic affinity of rockfowls (genus *Picathartes*) to crows and ravens (Corvidae), under review. Mol. Phylogenet. Evol.

Vandergon, T. L., Reitman, M., 1994. Evolution of chicken repeat-1 (CR-1) elements - evidence for ancient subfamilies and multiple progenitors. Mol. Biol. Evol. 11, 886-898.

Watanabe, M., Nikaido, M., Tsuda, T. T., Inoko, H., Mindell, D. P., Murata, K., Okada, N., 2006. The rise and fall of the CR1 subfamily in the lineage leading to penguins. Gene 365, 57-66.

Webster, M. T., Axelsson, E., Ellegren, H., 2006. Strong regional biases in nucleotide substitution in the chicken genome. Mol. Biol. Evol. 23, 1203-1216.

Wilhelm, M., Wilhelm, F. X., 2001. Reverse transcription of retroviruses and LTR retrotransposons. Cell. Mol. Life Sci. 58, 1246-1262.

Table 1 Primers for each CR1 locus used in this study.

Locus	Primers	Sequences	Annealing temperature
Darfin-CR1	CR1_Darfin_for	5' GAA CAG GTT GCC CAG AGA AGC TGT GGA TGT 3'	61°C
	CR1_Darfin_rev	5' GAG AAT GGA GGG GAA TCA GAG AAA TGG G 3'	
Fic1-CR1	CR1oW1	5' GCA CAG GTT GCC CAG AGA AGC TGT GG 3'	59°C
	Fic1rev	5' GCT GGG GGA CAG AGA GAG CTG GGG TT 3'	
Fic2-CR1	Fic2for	5' GCA CAC ACT GGA ACA GAT TGC C 3'	61°C
	Fic2rev	5' GCT GCC TGT CCT GTG TTT CTC A 3'	
Tur1-CR1	CR1oW1	5' GCA CAG GTT GCC CAG AGA AGC TGT GG 3'	55°C
	Tur1rev	5' GCC CTC TGC CTC TCC TCT GAA GTC T 3'	
Tur2-CR1	CR1oW1	5' GCA CAG GTT GCC CAG AGA AGC TGT GG 3'	58°C
	Tur2rev	5' GCA GCT GCC CCA CAT CTT AAC CCA 3'	
Par1-CR1	Par1for	5' GTT TTC CCT GTG AGA GTG GTG AG 3'	55°C
	Par1rev	5' GTG TTT TAA TTT TAC TTC CTG GTC TGC 3'	
Par2-CR1	Par2for	5' GAT TTC TTC ACT GTG AGG ATG G 3'	56°C
	Par2rev	5' GAG CAA GGT ACT GAG TCA TGT 3'	
Par3-CR1	Par3for	5' GAT TCA AAC TTA AAG AGA G 3'	50°C
	Par3rev	5' GTC TGC AGT GGA GGA TAT 3'	

Table 2 Presence of CR1 loci in passerine taxa and GenBank accession numbers (xxxxxxxx). Systematic classifications following Sibley and Monroe (1990). * Unsuccessful PCR-amplification. - Not evaluated. Darfin-CR1 sequences were used in both combined data sets MusCR1 and SylCR1.

Family - subfamily - tribe	species	Darfin-CR1	Fic1-CR1	Fic2-CR1	Tur1-CR1	Tur2-CR1	Par1-CR1	Par2-CR1	Par3-CR1
			MusCR1				SylCR1		
Sturnidae - Sturnini	*Sturnus vulgaris*	xxxxxxxx	xxxxxxxx	*	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	-
Muscicapidae - Turdinae	*Turdus merula*	xxxxxxxx	-	-	-	-	-	-	-
	Turdus philomelos	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	-	-	-
	Catharus guttatus	xxxxxxxx	-	xxxxxxxx	xxxxxxxx	xxxxxxxx	-	-	-
Muscicapidae - Muscicapinae - Muscicapini	*Muscicapa striata*	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	-	-	-
Muscicapidae - Muscicapinae - Saxicolini	*Erithacus rubecula*		xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	-	-	-
	Saxicola rubetra	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	-	-	-	-
	Luscinia luscinia	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	-	-	-
	Phoenicurus ochruros	xxxxxxxx	xxxxxxxx	*	xxxxxxxx	xxxxxxxx	-	-	-
	Phoenicurus phoenicurus	xxxxxxxx	xxxxxxxx	*	xxxxxxxx	xxxxxxxx	-	-	-
	Ficedula hypoleuca	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	xxxxxxxx	-	-	-
Bombycillidae	*Bombycilla garrulus*	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
Sittidae - Sittinae	*Sitta europaea*	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
Certhiidae - Certhiinae	*Certhia brachydactyla*	xxxxxxxx	-	-	-	-	xxxxxxxx	*	xxxxxxxx
Certhiidae - Troglodytinae	*Troglodytes troglodytes*	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	*
Paridae - Parinae	*Parus major*	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
	Parus caeruleus	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
	Parus cristatus	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
	Parus palustris	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
Hirundinidae - Hirundininae	*Delichon urbica*	xxxxxxxx	-	-	-	-	*	*	*
Regulidae	*Regulus regulus*	xxxxxxxx	-	-	-	-	*	xxxxxxxx	*
Pycnonotidae	*Pycnonotus leucogenys*	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
	Pycnonotus xanthopygos	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
Zosteropidae	*Zosterops senegalensis*	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	*
Sylviidae - Acrocephalinae	*Acrocephalus dumetorum*	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	*
	Hippolais icterina	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	*
	Phylloscopus trochilus	xxxxxxxx	-	-	-	-	*	xxxxxxxx	*

	Locustella naevia	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
Sylviidae - Garrulacinae	*Garrulax strepitans*	xxxxxxxx	-	-	-	-	xxxxxxxx	xxxxxxxx	xxxxxxxx
Sylviidae - Sylviinae - Sylviini	*Sylvia communis*	xxxxxxxx	-	-	-	-	*	xxxxxxxx	xxxxxxxx
Alaudidae	*Alauda arvensis*	xxxxxxxx	-	-	-	-	*	xxxxxxxx	xxxxxxxx

Table 3 Summary statistics of sequences and MP trees for all loci and the combined loci analyses.

	Darfin1 (Mus)	Fic1	Fic2	Tur1	Tur2	MusCR1	Darfin (Syl)	Par1	Par2	Par3	SylCR1
Size (bp)	148	155	155	147	137	742	152	186	183	221	742
PI[a]	10	18	18	12	14	70	43	40	36	32	151
%G	24.4	30.0	30.0	28.3	25.3	27.4	24.7	27.8	27.3	28.4	27.5
%A	22.3	20.7	18.8	22.0	25.0	21.8	22.0	22.1	24.7	25.0	23.2
%T	27.2	25.5	24.4	26.9	27.4	26.5	28.3	26.5	26.8	24.3	26.8
%C	26.1	23.7	26.8	22.8	22.2	24.3	25.0	23.7	21.1	22.3	22.5
Ts/tv ratio	1.442	3.133	2.938	1.870	2.815	2.197	1.335	1.290	1.702	2.095	1.596
Model[b]	K80	K80+I	K80	K80	K80+I	GTR+G	HKY+G	K80	K80	K80	HKY+G
PINVAR	0.000	0.357	0.000	0.000	0.517	0.000	0.000	0.000	0.000	0.000	0.000
Gamma	-	-	-	-	-	1.180	1.541	-	-	-	2.044
minimum divergence	see text	0.007	0.023	0.021	0.000	not computed	see text	0.000	0.000	0.000	not computed
maximum divergence	see text	0.156	0.171	0.141	0.113	not computed	see text	0.207	0.204	0.166	not computed
CI[c]	0.938	0.870	0.868	0.868	0.833	0.854	0.713	0.854	0.818	0.881	0.781
RI[c]	0.824	0.767	0.759	0.650	0.708	0.670	0.634	0.737	0.654	0.809	0.638
RC[c]	0.772	0.667	0.658	0.564	0.590	0.572	0.452	0.629	0.535	0.712	0.498

[a] Parsimonious informative sites.

[b] Models of molecular evolution represent the Kimura80 (K80=K2P) model (Kimura 1980), general time-reversible (GTR) model (Tavaré *et al*. 1986), and the Hasegawa-Kishuno-Yano (HKY) model (Hasegawa *et al*. 1985) with assumptions of proportions of invariable sites (PINVAR) or gamma shape correction parameters (Page and Holmes 1998, Swofford, 2001).

[c] Measures of homoplasy (CI, RI, and RC values) are given for *n* equally parsimonious trees.

Table 4a *p*-values of the heterogeneity test for all combinations of the Mus-CR1 loci.

	Darfin-CR1	Fic1-CR1	Fic2-CR1	Tur1-CR1
Fic1-CR1	0.404			
Fic2-CR1	0.868	0.384		
Tur1-CR1	0.990	0.357	0.996	
Tur2-CR1	1.000	0.229	1.000	0.949

Table 4b *p*-values of the heterogeneity test for all combinations of the Syl-CR1 loci.

	Darfin-CR1	Par1-CR1	Par2-CR1
Par1-CR1	0.956		
Par2-CR1	0.653	0.544	
Par3-CR1	0.999	0.679	1.000

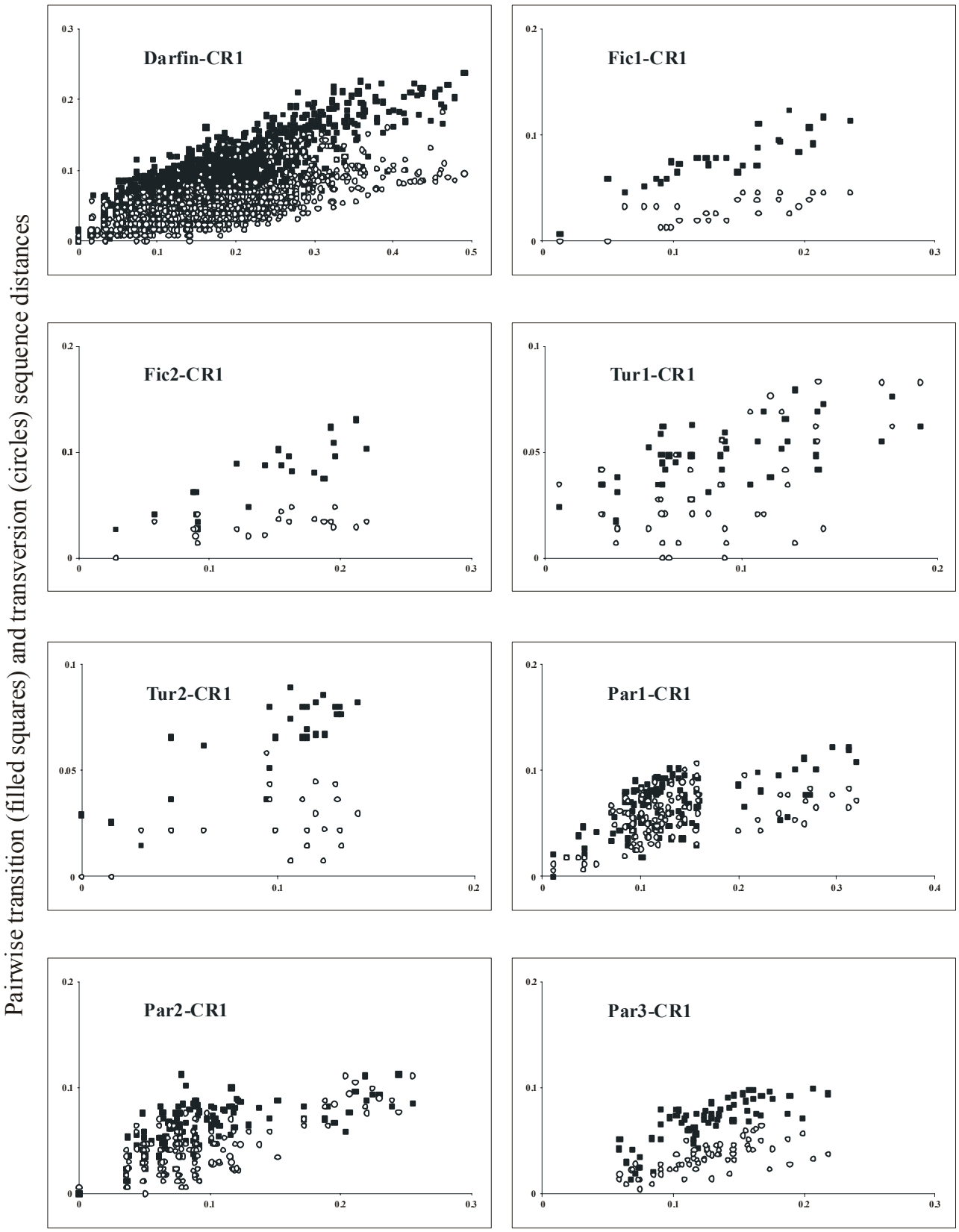
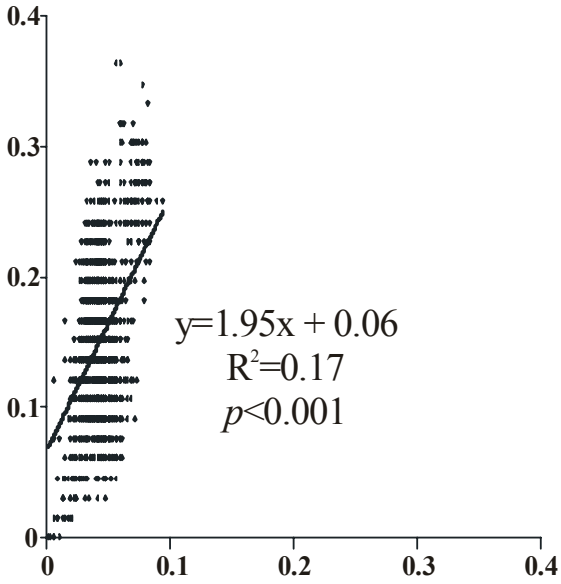
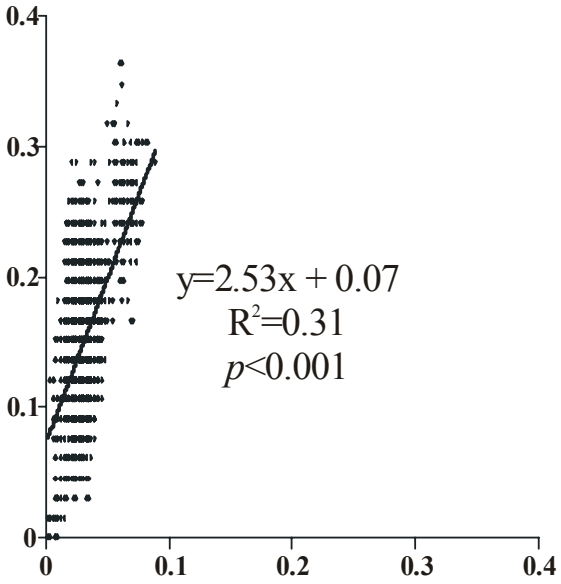
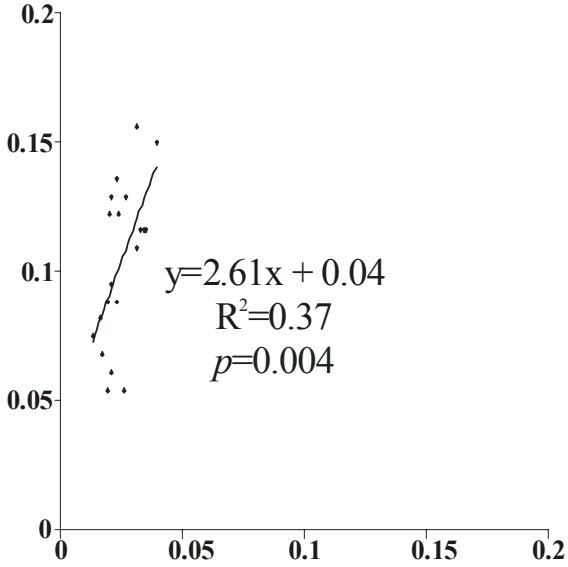


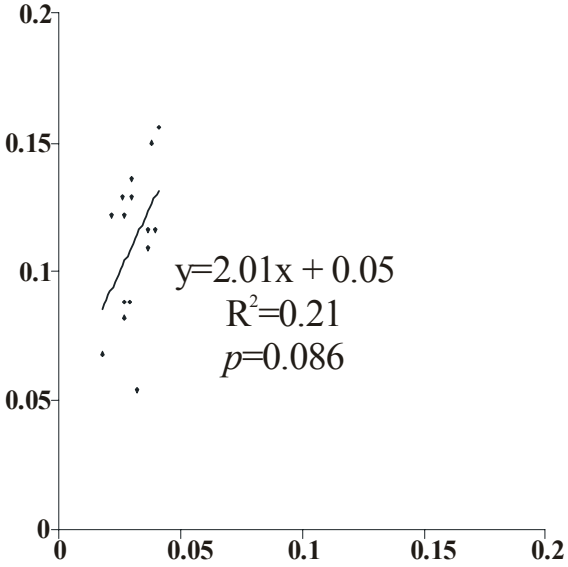
Fig. 1 Saturation plots. Pairwise transition and transversion sequence distance plotted against total sequence divergence for each CR1 locus.



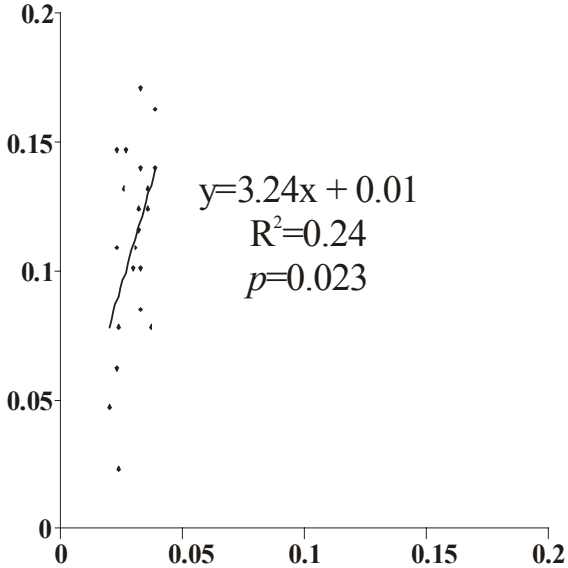


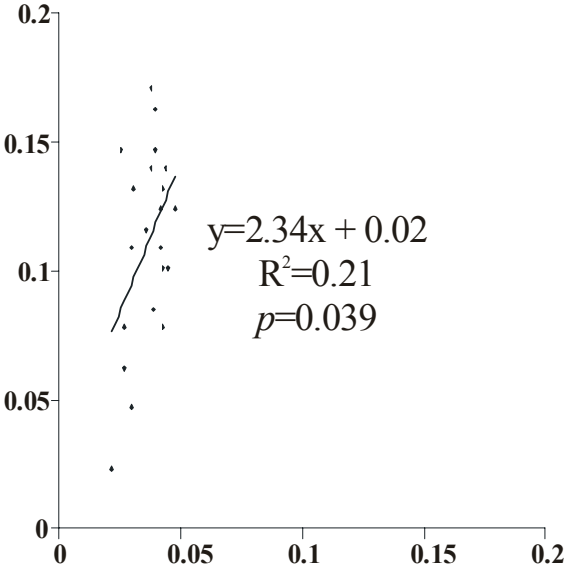
a. Darfin-CR1



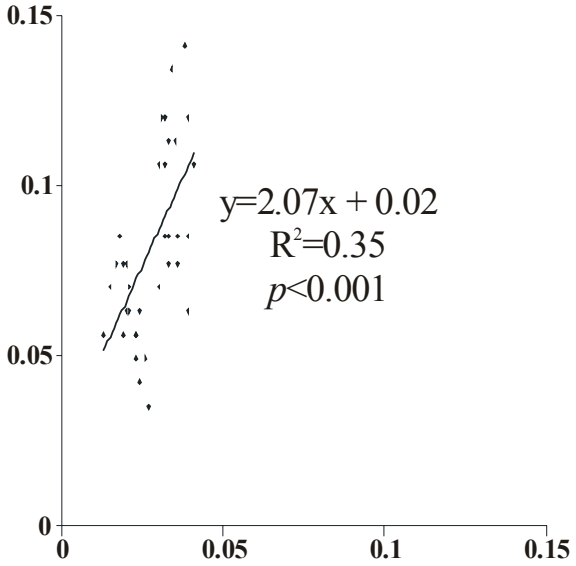


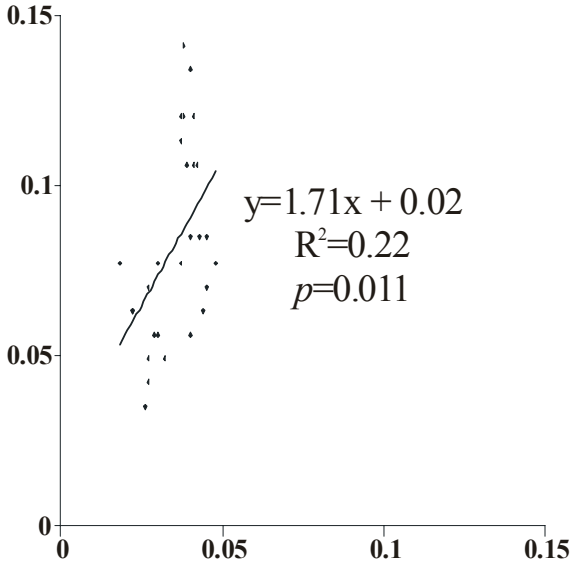
b. Fic1-CR1



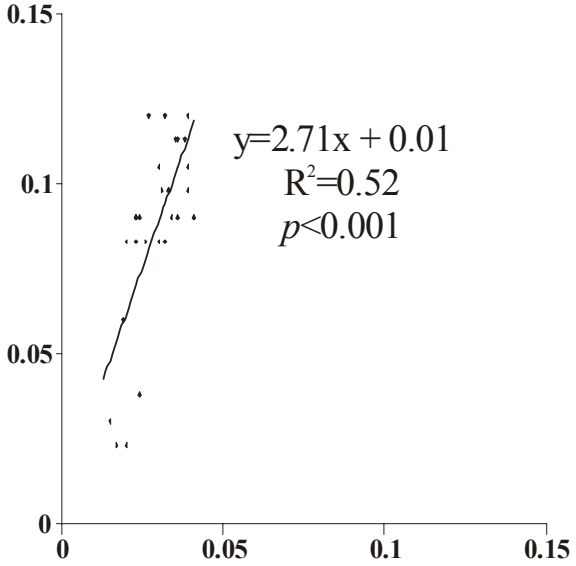


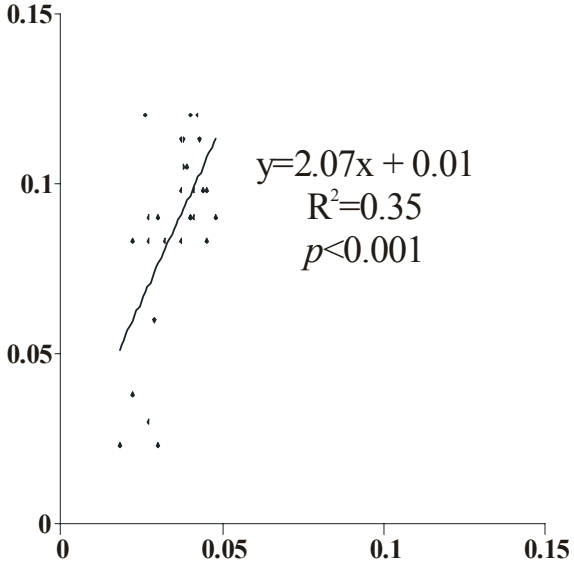
c. Fic2-CR1



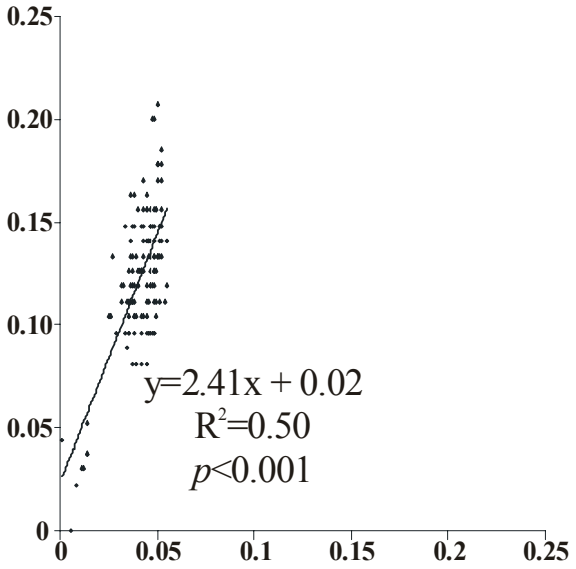


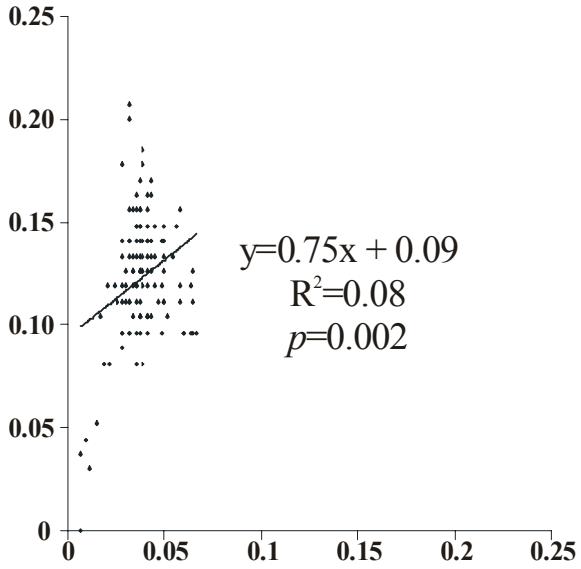
d. Tur1-CR1



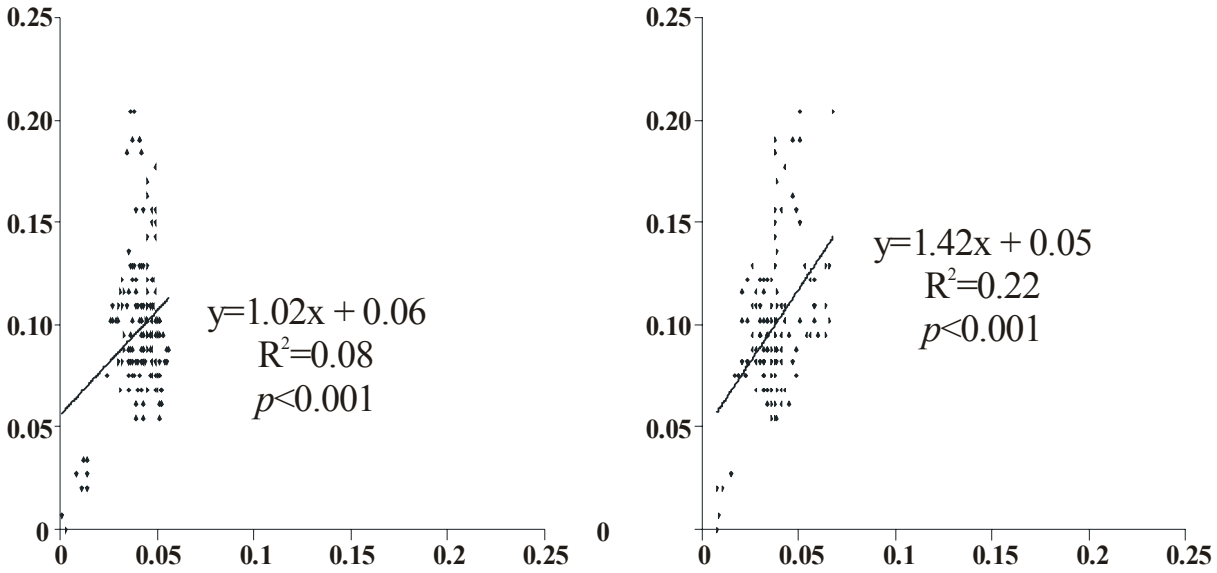


e. Tur2-CR1

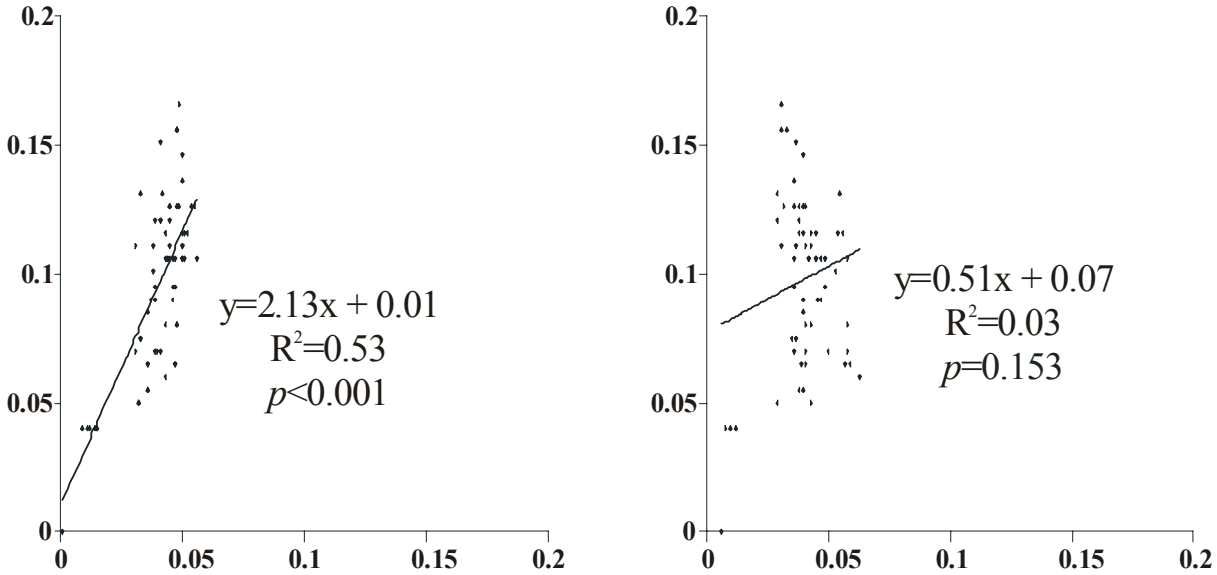




f. Par1-CR1



g. Par2-CR1



h. Par3-CR1

Fig. 2 a-h Total distances of CR1 loci plotted against total distances of the nuclear genes ZENK (left) and RAG-1 (right). *p*-values indicate significance of correlation between the nuclear markers.

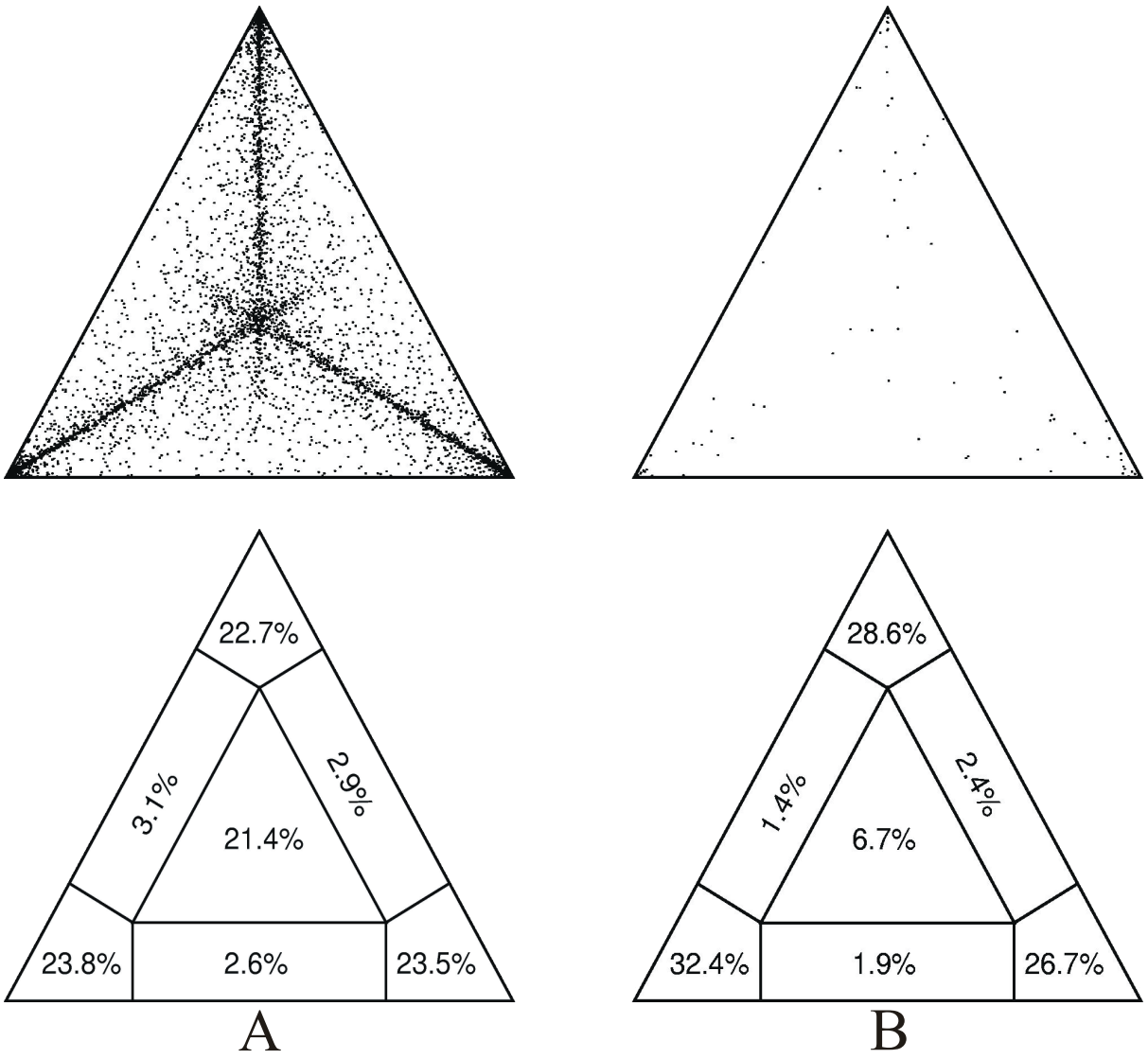


Fig. 3 Likelihood-mapping analyses of Syl-CR1 (A) and Mus-CR1 (B) data sets with distribution patterns (upper triangles) and percentages of the seven areas of attraction (lower triangles).

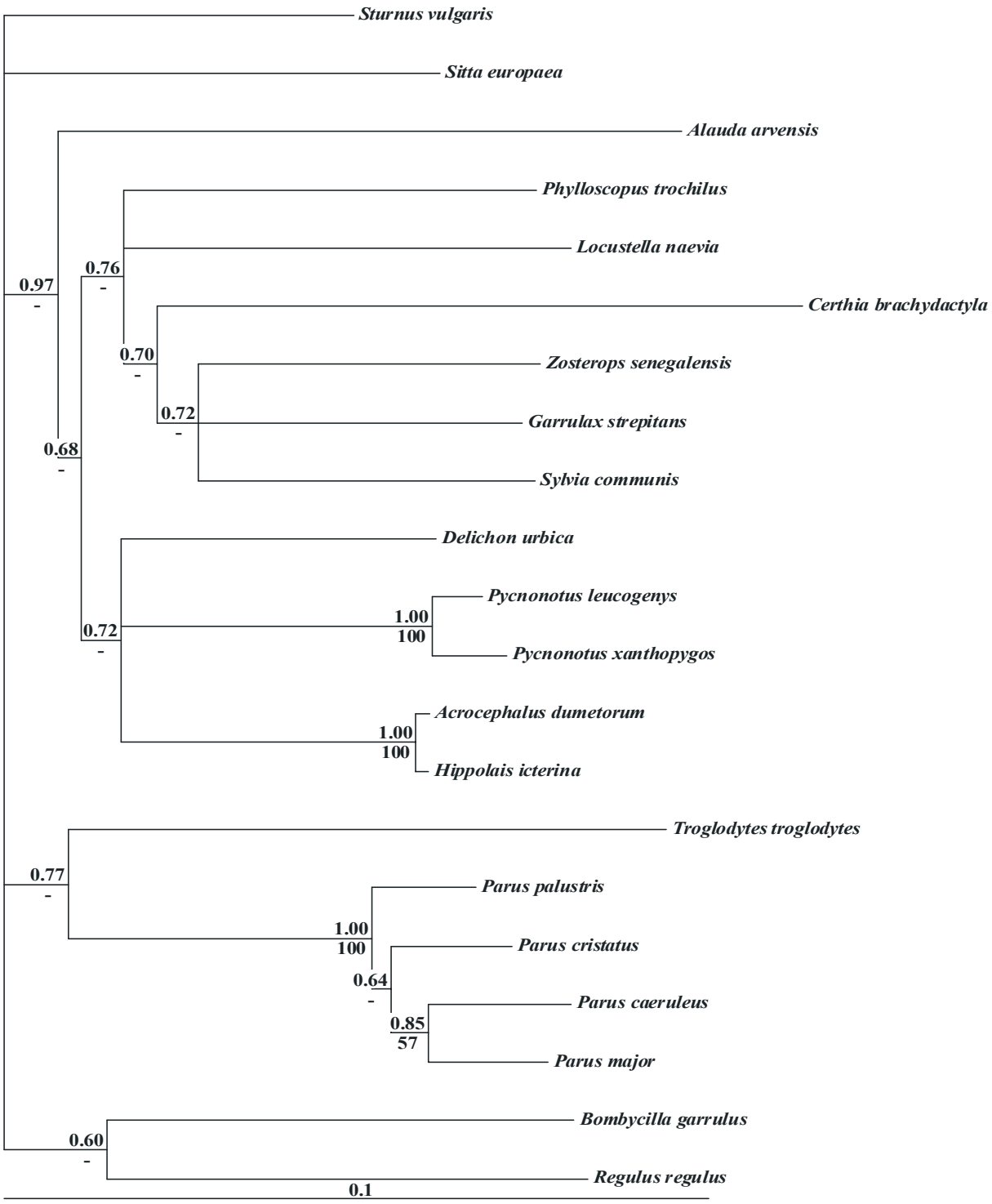


Fig. 4 Phylogenetic tree of the Bayesian analysis of the combined Syl-CR1 data set. Bayesian support values are given above, MP bootstrap support below the nodes. Hyphens indicate unresolved nodes in the MP bootstrap analysis.

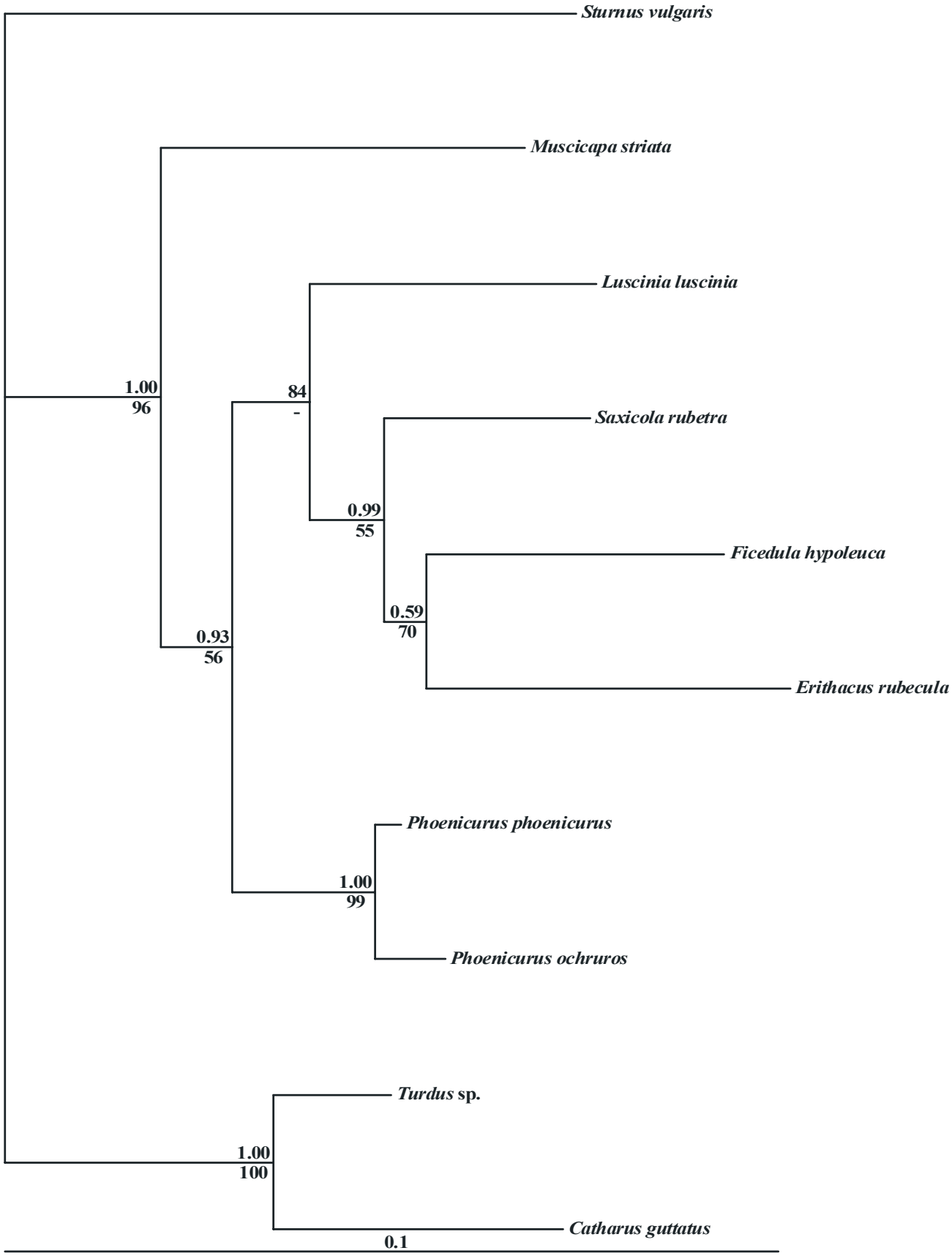


Fig. 5 Phylogenetic tree of the Bayesian analysis of the combined Mus-CR1 data set. Bayesian support values are given above, MP bootstrap support below the nodes.