

Accepted Manuscript

Comprehensive molecular phylogeny of the grassbirds and allies (Locustellidae) reveals extensive non-monophyly of traditional genera, and a proposal for a new classification

Per Alström, Alice Cibois, Martin Irestedt, Dario Zuccon, Magnus Gelang, Jon Fjeldså, Michael J. Andersen, Robert G. Moyle, Eric Pasquet, Urban Olsson

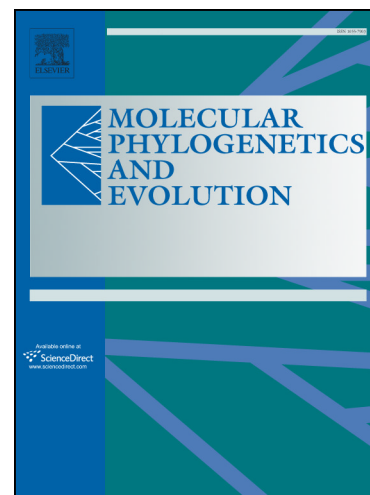
PII: S1055-7903(17)30719-4
DOI: <https://doi.org/10.1016/j.ympev.2018.03.029>
Reference: YMPEV 6101

To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 5 October 2017
Revised Date: 21 February 2018
Accepted Date: 28 March 2018

Please cite this article as: Alström, P., Cibois, A., Irestedt, M., Zuccon, D., Gelang, M., Fjeldså, J., Andersen, M.J., Moyle, R.G., Pasquet, E., Olsson, U., Comprehensive molecular phylogeny of the grassbirds and allies (Locustellidae) reveals extensive non-monophyly of traditional genera, and a proposal for a new classification, *Molecular Phylogenetics and Evolution* (2018), doi: <https://doi.org/10.1016/j.ympev.2018.03.029>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



**Comprehensive molecular phylogeny of the grassbirds and allies
(Locustellidae) reveals extensive non-monophyly of traditional genera, and
a proposal for a new classification**

Per Alström^{1,2,3,*}, Alice Cibois⁴, Martin Irestedt⁵, Dario Zuccon⁶, Magnus Gelang⁷, Jon Fjeldså⁸, Michael J. Andersen⁹, Robert G. Moyle¹⁰, Eric Pasquet⁶, Urban Olsson¹¹

¹ Department of Ecology and Genetics, Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36, Uppsala, Sweden

² Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, Uppsala SE-750 07, Sweden

³ Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

⁴ Department of Mammalogy and Ornithology, Natural History Museum of Geneva, CP 6434, CH 1211 Geneva 6, Switzerland

⁵ Department of Bioinformatics and Genetics, Swedish Museum of Natural History, PO Box 50007, Stockholm, SE-10405, Sweden

⁶ UMS MNHN/CNRS 2700 Outils et Méthodes de la Systématique Intégrative (OMSI) & UMR7205 Institut de Systématique, Evolution, Biodiversité CNRS MNHN UPMC EPHE, Sorbonne Universités, Muséum National d'Histoire Naturelle, CP 51, 57 rue Cuvier, F-75231 Paris Cedex 05, France

⁷ Gothenburg Natural History Museum, Box 7283, 402 35 Göteborg, Sweden

⁸ Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, Zoological Museum, Universitetsparken 15, DK-2100 Copenhagen, Denmark

⁹ Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico, 87131, USA

¹⁰ Department of Ecology and Evolutionary Biology and Biodiversity Institute, University of Kansas, Lawrence, Kansas, 66045, USA

¹¹ University of Gothenburg, Department of Biology and Environmental Sciences, Systematics and Biodiversity, Box 463, 405 30 Göteborg, Sweden

* Corresponding author: Per Alström, Department of Ecology and Genetics, Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36, Uppsala, Sweden. Email: per.alstrom@ebc.uu.se

Abstract

The widespread Old World avian family Locustellidae ('grassbirds and allies') comprises 62 extant species in 11 genera. In the present study, we used one mitochondrial and, for most species, four nuclear loci to infer the phylogeny of this family. We analysed 59 species, including the five previously unsampled genera plus two genera that had not before been analysed in a densely sampled dataset. This study revealed extensive disagreement with current taxonomy; the genera *Bradypterus*, *Locustella*, *Megalurus*, *Megalurulus* and *Schoenicola* were all found to be non-monophyletic. Non-monophyly was particularly pronounced for *Megalurus*, which was widely scattered across the tree. Three of the five monotypic genera (*Amphilais*, *Buettikoferella* and *Malia*) were nested within other genera; one monotypic genus (*Chaetornis*) formed a clade with one of the two species of *Schoenicola*; whereas the position of the fifth monotypic genus (*Elaphrornis*) was unresolved. *Robsonius* was confirmed as sister to the other genera. We propose a phylogenetically informed revision of genus-level taxonomy, including one new generic name. Finally, we highlight several non-monophyletic species complexes and deep intra-species divergences that point to conflict in taxonomy and suggest an underestimation of current species diversity in this group.

Keywords: systematics; taxonomy; incongruence; species diversity; new genus

1. Introduction

Sylvioid songbirds (Sylvioidea sensu Fregin et al., 2012) include for example all Old World 'warblers' (several families), 'babblers' (several families), swallows (Hirundinidae), bulbuls (Pycnonotidae) and larks (Alaudidae) (review in Alström et al. 2013a). One of the 'warbler' families in this assemblage is the Locustellidae ('grassbirds and allies'). This family has erroneously been referred to as Megaluridae (e.g. by Alström et al., 2006; Johansson et al., 2008), but Locustellidae has priority (Bock, 1994: p. 152). This family consists of 62 extant and one recently extinct species (Gill and Donsker, 2017), which are widely distributed across Africa, Eurasia and Oceania. The family has a chequered taxonomic history, both at the generic and species level (review in Alström et al., 2013a).

Only one broad phylogenetic analysis has been published (Alström et al., 2011a), which revealed several non-monophyletic genera, and proposed a taxonomic revision. For instance,

the Asian *Bradypterus* were synonymised with *Locustella* and *Dromaeocercus* with *Bradypterus* (hence restricting the latter to African and Malagasy species), and *Eremiornis* and *Cincloramphus* were synonymised with *Megalurus*. However, these authors stressed that the circumscription of *Megalurus* was preliminary, as the type species (*M. palustris*) was sister to the *Bradypterus sensu stricto* clade in their analysis, albeit with low support. In contrast, Dickinson and Christidis (2014) and del Hoyo and Collar (2016) split *Megalurus* into three genera (*Megalurus sensu stricto*, *Poodytes* and *Cincloramphus*) based on the results of Alström et al. (2011a). Oliveros et al. (2012) unexpectedly found *Robsonius* and *Malia* to be part of Locustellidae.

Five genera traditionally placed in Locustellidae have not been analysed phylogenetically: *Amphilais* (monotypic, Madagascar), *Megalurulus* (six species, Melanesia), *Buettikoferella* (monotypic, Timor), *Chaetornis* (monotypic, Indian Subcontinent) and *Elaphrornis* (monotypic, Sri Lanka) (Gill and Donsker, 2017). Here we reconstruct the phylogeny of Locustellidae, including all genera and 59 of 62 recognised species (Gill and Donsker, 2017) using mitochondrial and nuclear loci. We present a chronogram, which includes multiple subspecies for some polytypic species, and propose a revised taxonomic classification of the Locustellidae.

2. Material and methods

2.1. Study group

We analysed sequence data from 122 individuals of 59 extant and one recently extinct (*Megalurus rufescens*) species, representing all genera and all but three species (Gill and Donsker, 2017; Supplementary Table S1). As outgroups, we used representatives of the presumed most closely related families, Bernieridae, Donacobiidae and Acrocephalidae (Fregin et al., 2012; Supplementary Table S1).

2.2. Lab work

DNA was extracted from fresh material using the Qiagen DNA Mini Kit and following the manufacturer's protocol. Twenty three toepad samples (17 species) were obtained from museum specimens (Supplementary Table S1), and most of these were extracted in clean aDNA-dedicated spaces, using the Qiagen DNA Micro Kit and the protocol detailed in Irestedt et al. (2006). We sequenced the mitochondrial cytochrome *b* (*cytb*) gene and four nuclear regions: myoglobin intron 2 (*myo*), ornithine decarboxylase (mainly) introns 6–7

(ODC), glyceraldehyde-3-phosphodehydrogenase intron 11 (GAPDH) and lactate dehydrogenase intron 3 (LDH). Amplification and sequencing of the fresh samples followed the protocols described in Fregin *et al.* (2012). The toepads were sequenced in short (150–300 bp) segments with specifically designed primers and specific amplification profiles (Supplementary Table S2). Not all loci were obtained for all species, and for eight species only *cytb* was available (Supplementary Table S1). In addition, we downloaded sequences from GenBank of the recombination-activating gene 1 (RAG1) for the 9 Locustellidae species for which this gene was available as well as 9 outgroup species (GenBank numbers in Supplementary Fig. S4).

Authenticity of sequences obtained from toepad samples is supported by several lines of evidence. (1) When independent samples from the same species were included, the sequences were always highly similar. (2) Phylogenetic relationships based on individual PCR amplicons were the same as those using full *cytb* contigs. (3) No fragment was identical to any other species included in this study. (4) Overlapping forward and reverse sequence fragments were identical. (5) The mitochondrial sequences showed no double signal in the electropherograms or stop codons, insertions or deletions, and a vast majority of nucleotide substitutions were found in the 3rd codon position and resulted in few amino acid substitutions (of which a majority also was found in sequences obtained from the fresh samples). The mitochondrial sequences from fresh samples were also validated in the same way.

2.3. Phylogenetic analyses

Sequences were aligned and checked using Geneious 7.1.9 (Biomatters Ltd.). For the nuclear loci, heterozygous sites were coded as ambiguous. Trees were estimated by Bayesian inference using BEAST 1.8.4 (Drummond *et al.* 2012) with different data partitioning schemes: (1) all loci were analysed separately (single-locus analyses, SLAs); (2) all sequences except RAG1 were concatenated and partitioned by locus (RAG1 excluded because only few species were available); and (3) nuclear loci except RAG1 were concatenated and partitioned by locus.

All analyses were run under the best-fit models according to the Bayesian Information Criterion calculated in jModeltest 2.1.7 (Darriba *et al.* 2012). The following models were selected: *cytb*, GTR + Γ + I; *myo*, HKY; GAPDH, LDH and RAG1, HKY + Γ ; and ODC, HKY + I. An uncorrelated relaxed clock model with a lognormal distribution was applied to all partitions. Substitution models and clock models were unlinked. A ‘birth-death incomplete sampling’ tree prior was applied. Default priors were used except for the

'birthDeath.meanGrowthRate', for which a normal prior with an initial value 1.0, mean 2.0 and standard deviation 1.0 was applied. Xml files were generated in the BEAST utility program BEAUti version 1.8.4. The analyses were run for 50–100 million generations and sampled every 1000 generations, and each analysis was run twice.

Integrative species tree estimation was performed using *BEAST (Heled and Drummond, 2010) in BEAST 1.8.4, with gene trees and species trees estimated simultaneously. We ran analyses under the same substitution models per partition as in the previous analyses, and an uncorrelated lognormal relaxed clock prior (Drummond *et al.*, 2006). A piecewise linear population size model with a constant root was used as a prior for the multispecies coalescent and a birth-death model as prior on divergence times. Default settings were used for the priors. 100–150 million generations were run in different runs, sampled every 1000 generations; the analysis was repeated four times.

In order to estimate divergence times and intraspecific variation, the *cytb* data set with multiple subspecies was analysed in BEAST version 1.8.4 (Drummond *et al.* 2012). Analyses were run under the GTR + Γ model (cf. Weir and Schluter, 2008) with a 'birth-death incomplete sampling' species tree prior with a normal distribution with mean 2.0 and standard deviation 1.0. A strict clock with a mean rate of 2.1%/million years (Weir and Schluter, 2008) and a normal prior distribution with standard deviation 0.001 was applied. Default settings were used for the other priors. 100 million generations were run, sampled every 1000 generations. The analysis was run twice. Nodes B and I were constrained in the final analysis based on the results from the multilocus analyses (cf. Fig. 1), as these clades were not supported by *cytb* alone (and no alternative topology was strongly supported in the unconstrained *cytb* tree).

In all BEAST and *BEAST analyses, convergence to the stationary distribution of the single chains was inspected in Tracer 1.6 (Rambaut *et al.*, 2014). The effective sample sizes (ESS) for the joint likelihood and other parameter values were >1000, representing good mixing of the MCMC, except in the *BEAST analyses, where ESSs were 100–150. We also examined convergence and reproducibility by running each analysis at least twice, with random starting points, and comparing the results. In all analyses, including the *BEAST analyses with low ESSs, the topologies (including relative branch lengths) and posterior probabilities (PPs) were similar across different runs. The first 25% of generations were discarded as 'burn-in', and the PPs were calculated from the remaining samples. Samples were combined in LogCombiner 1.8.4, and trees were summarized using TreeAnnotator version 1.8.4 (both included in the BEAST package), choosing 'Maximum clade credibility

tree' and 'Mean heights', and displayed in FigTree version 1.4.3 (Rambaut 2002). Xml files for all multilocus analyses are available as Supplementary Material 1.

The concatenated sequences (except RAG1) partitioned by locus were also analysed by Maximum Likelihood bootstrapping (MLBS). MLBS (1000 replicates) was conducted with RAxML-HPC2 version 8.2.10 (Stamatakis, 2014). Default parameters were used.

Because *Robsonius* was sister to the rest of Locustellidae in all preliminary analyses, we tested that this genus is indeed part of this clade and not more distantly related by analysing a dataset including representatives from the primary clades within the Passerida group as revealed by previous studies (e.g. Alström *et al.* 2006; Fregin *et al.* 2012; Alström *et al.* 2014; Moyle *et al.* 2016), using sequences from Alström *et al.* (2014). This was run in MrBayes 3.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) using concatenated sequences partitioned by locus and the following models: *cytb* and ODC GTR + Γ + I; *myo* and GAPDH HKY + Γ ; and LDH GTR + Γ . Default priors in MrBayes were used. Four Metropolis-coupled MCMC chains were run for 5 million generations and sampled every 1000 generations. Convergence was checked as for the BEAST analyses, as well as by the average standard deviation of split frequencies passing below 0.01 and the potential scale reduction factor (PSRF) being close to 1.00 for all parameters.

Several of the BI and the MLBS analyses were run on the CIPRES Science Gateway (Miller *et al.*, 2010).

2.5 Song comparisons

Sound recordings of territorial songs were obtained from all *Locustella* species (own recordings and from xeno-canto [www.xeno-canto.org] and British Library National Sound Archive). Sonograms were generated in Raven Pro 1.5 (Cornell Laboratory of Ornithology, Ithaca, USA) to graphically illustrate differences among species. All of our own sound recordings used for this study have been deposited in xeno-canto (www.xeno-canto.org), and detailed information is available as available as Supplementary Material 2.

3. Results

3.1. Species-level phylogeny

The *BEAST phylogeny based on all sequences except RAG1 (Fig. 1) identified *Robsonius* as sister to all other locustellid species, which were divided into two primary clades (A and B). Clade A (PP 1.00) comprised all genera except *Locustella*, and clade B (PP

0.94) contained all *Locustella* and one *Bradypterus*. Within clade A, five strongly supported (PP 1.00) clades were found (C–G), each containing 1–4 genera. Clades C–E plus *Schoenicola brevirostris* and *Elaphrornis palliseri* formed the poorly supported (PP 0.84) clade H; and clades F–G formed clade I (PP 0.94). Clade H comprised six genera from the Afrotropical, Madagascan and Oriental regions, and clade I included four genera mainly distributed in the Indo-Pacific. Within clade B, two main clades (J, K) were strongly supported (PP 1.00). All of the polytypic genera were non-monophyletic. The Afrotropical *Bradypterus* clade (C) also contained the monotypic genus *Amphilais* from Madagascar, whereas the Afrotropical *Bradypterus alfredi* was in one of the *Locustella* clades (J). *Megalurus* and *Megalurulus* were scattered across clades D, F and G, and the monotypic *Buettikoferella* and *Malia* were also nested among these. Finally, the two *Schoenicola* species were not sisters. Instead, the Oriental *S. platyurus* was strongly supported as sister to the monotypic *Chaetornis*, whereas the position of the Afrotropical *S. brevirostris* was uncertain.

The topology of the BEAST tree based on the concatenated sequences (except RAG1) was mostly in agreement with the *BEAST tree (Supplementary Fig. S1). The former found stronger support for clade B (1.00 vs. 0.94) and I (0.98 vs. 0.94) as well as a number of the relationships within smaller clades (highlighted in blue in Supplementary Fig. S1). The position of *Robsonius* as sister to the rest of Locustellidae was confirmed in the analysis of a broader sampling from Passeriformes (Supplementary Fig. S2). The tree based on concatenated nuclear loci (except RAG1; Supplementary Fig. S3) strongly supported relationships among clades J1–J5 that were incongruent with those in the *BEAST (Fig. 1) and *cytb* (Fig. 2) trees. Single-locus analyses, especially of the nuclear loci, were generally not well resolved (Supplementary Figs S4 and S5). Maximum Likelihood bootstrapping of the complete dataset (except RAG1) was generally in agreement with the BI analyses, especially the BEAST analysis of concatenated sequences (Fig. 1).

3.2. Dating and intraspecific variation

The chronogram based on *cytb* and containing subspecies of some species (Fig. 2) has generally wide confidence intervals. Our dating suggests that *Robsonius* diverged from the rest of Locustellidae at 22.7 mya (95% HPD 16.9–29.2 mya), and the age of the split between clades A and B was estimated at 15.4 mya (95% HPD 12.1–19.0 mya). The two primary clades A and B were estimated to have diversified since around 14.5 mya (95% HPD c. 11–18 mya). Six of the main clades (C, E, F, G, J, K) diversified during the period c. 7.8–11.4 my (95% HPD c. 5.6–14.3 mya).

Several of the polytypic species included subspecies estimated to have separated >2 mya: *Bradypterus lopezi ufipae* vs. *B. l. mariae* (2.1 mya; 95% HPD 1.3–3.0 mya); *Megalurus palustris forbesi* vs. *M. p. toklao* (4.0 mya; 2.7–5.5 mya); and *Megalurus punctatus caudatus* vs. *M. p. vealeae* (2.2 mya; 1.4–3.1 mya). Also the two samples of *Locustella caudata unicolor* were deeply diverged (2.7 mya; 1.8–3.7 mya), as were the two samples of the monotypic *Bradypterus brunneus* (5.7 mya; 4.1–7.5 mya). In addition, *Megalurus timoriensis* and *M. macrurus* were inferred to be non-monophyletic, with deep divergences between *M. timoriensis tweedalei*–*M. t. crex* and *M. macrurus macrurus*–*M. m. interscapularis*, respectively. The non-sisters *Locustella ochotensis*–*L. certhiola* were not monophyletic, as one individual of the former was nested within the latter.

3.3. Morphological, vocal and behavioural differences between species in clades J and K

Within *Locustella* (clade B), the species in clades J and K differ in that the latter are generally larger (clade K: 13–18 cm, 12–33 g; clade J: 12–18 cm, 9–21 g; del Hoyo et al. 2006). The lower mandible is always pale in the species in clade K, whereas it is all black in the breeding season, at least in males, in many of the species in clade J (Kennerley and Pearson 2010; pers. obs.).

The songs of the species in clade K consist of short (c. 2–5 s) strophes separated by distinct pauses (c. 2–15 s; shorter strophes when excited; Fig. 3). All or most of the elements in the strophes are either different from each other (in *L. amnicola*, *L. fasciolata*, *L. pryeri*) or arranged in a few to several different “blocks” of similar notes (in *L. certhiola*, *L. ochotensis*, *L. pryeri*). In contrast, the songs of the species in clade J consist of short, comparatively simple syllables, which are monotonously repeated at very short intervals (e.g. in *L. accentor*, *L. thoracica*, *L. davidi*) or in drawn-out, rattling reels (e.g. in *L. naevia*, *L. luteoventris*; Fig. 3). The songs of the species in clade J may continue without any distinct pauses for up to a few minutes. The song of *L. pryeri* in clade K may seem rather similar to that of e.g. *L. tacsanowskia* in clade J, but the song elements are more varied, without the regularly repeated pattern of *L. tacsanowskia* and the other species in clade J (Fig. 3). In all species in clade K except *L. fasciolata* and *L. amnicola*, the song is regularly delivered in a short song-flight, unlike in the species in clade J (Fig. 3; Kennerley and Pearson 2010; pers. obs.).

4. Discussion

4.1. Phylogeny

Our study is the most comprehensive analysis of the Locustellidae. It includes five previously unsampled monotypic genera and in total 11 previously unstudied species and three additional species that have not been analysed in a study with a large number of species. It also includes one extinct species, *Megalurus rufescens*. Only *Bradypterus grandis*, *Megalurus albolimbatus* and *Megalurulus llanae* were not analysed.

The majority of nodes (~60%) are well supported (PP \geq 0.95). However, some deep relationships remain uncertain. In particular, the support is poor for clade H, and the relationships among clades C, D, E, *Schoenicola brevirostris* and *Elaphrornis palliseri* are equivocal, likely due to their short internode distances. We note that the support for the position of *Schoenicola brevirostris* as sister to clade C is stronger in the concatenation analysis of all loci (PP 1.00; MLBS 83%) and in the concatenation analysis of the nuclear loci (PP 1.00), and the support for clade I is stronger in the concatenation analyses of all loci (PP 0.97; MLBS 78%). This makes sense from a biogeographical point of view – *S. brevirostris* is an Afrotropical species, like all *Bradypterus*.

The lack of evidence for a close relationship between *Schoenicola brevirostris* and *S. platyurus* is surprising, because they have been treated as conspecific due to their very similar appearances (e.g. Watson et al., 1986), while they differ from *Bradypterus* in for example their broader and more strongly graduated tails with pale tips (Madge, 2017). In contrast, *Chaetornis striata*, which we found as sister to *S. platyurus*, is morphologically markedly different, although it shares the pale-tipped rectrices with the two *Schoenicola* species, and it occurs in the same geographical area as *S. platyurus* (Indian subcontinent) (Madge, 2017). We hypothesise that *S. brevirostris* is sister to clade E (as suggested by *cytb*, although with no support), and that the plumage similarity between *S. brevirostris* and *S. platyurus* is plesiomorphic.

The Sri Lankan endemic *Elaphrornis palliseri* was previously placed in *Bradypterus* (e.g. Watson et al., 1986), and its move to *Elaphrornis* was only based on it being “entirely distinct from the genus *Bradypterus*” (Dickinson, 2003). The monotypic *Elaphrornis* is here shown to be distinct, although a close relationship with the *Bradypterus* clade (C) cannot be excluded. In contrast, the Malagasy endemic *Amphilais seebohmi* is firmly nested in *Bradypterus*, as sister to the Malagasy endemic *B. brunneus*. *Amphilais seebohmi* is sometimes placed in the genus *Dromaeocercus*, together with *B. brunneus* (e.g. Watson et al., 1986), and is treated as *incertae sedis* by Dickinson and Christidis (2014).

The position of *Megalurus palustris* in clade H is not strongly supported in any of our multilocus analyses, and the only SLA that strongly corroborates this is *cytb*, whereas

analysis of a small set of RAG1 sequences (Supplementary Fig. S5) strongly supports a sister relationship between *M. palustris* and *M. mathewsi* (which is the only representative of clade I in that analysis). The same was found by Alström et al. (2011a), who discussed this at length, and concluded that inclusion of *Megalurus palustris* in the equivalent of clade H is surprising from both a morphological, vocal and biogeographical point of view. If *M. palustris* indeed belongs in clade H, the most parsimonious position would be as sister to the others, as the strong similarities in plumage, structure, size and voice to some of the other species in clade I could then be explained as being plesiomorphic.

The non-monophyly of *Megalurus* is further exaggerated within clade I, as the two monotypic genera *Malia* and *Buettikoferella* as well as the five species of *Megalurulus* are intermixed with *Megalurus*. Alström et al. (2011a) also found the equivalent of this clade to contain a non-monophyletic *Megalurus* (see below), although they did not include *Malia*, *Buettikoferella* or *Megalurulus*. Oliveros et al. (2012) recently discovered that *Malia grata*, with previously unknown affinities, is part of Locustellidae; it was recovered with strong support as sister to *Megalurus timoriensis*, which was the only representative for our clade I in their study. *Buettikoferella bivittata* is firmly anchored in a clade with *Megalurus mathewsi*, *M. timoriensis*, *M. macrurus* and the five *Megalurulus* species, although its sister relationship with *M. mathewsi* is poorly supported. Also *Megalurulus* is not monophyletic; none of the relationships among the species are strongly supported by *BEAST, although the relationships among *M. rufus*, *M. mariae* and *M. whitneyi* are well supported in the BI and ML analyses of concatenated data.

The strong non-monophyly of *Megalurus* and *Megalurulus* is unexpected. Except for *M. carteri* and male *M. cruralis*, the members of the genus *Megalurus*, including *M. palustris*, form a morphologically fairly homogeneous group (del Hoyo et al., 2006; cf. Fig. 1). From a morphological perspective, the species in *Megalurulus* and *Buettikoferella bivittata* are fairly similar, while they differ much from *Megalurus* by their more uniform and more saturated brown plumage colorations (del Hoyo et al., 2006; cf. Fig. 1). The differences appear to be adaptive: *Megalurulus* species occur in understory of evergreen mesic habitats, whereas *Megalurus* species are associated with grasslands or more arid scrubby habitats. *Malia grata* is highly aberrant in plumage and ecology (mainly arboreal, social; Collar and Robson, 2017; pers. obs.) compared to all other members of Locustellidae (cf. Fig. 1). Strongly divergent plumages, such as the green and yellow plumage of *Malia grata* and the boldly patterned male *Megalurus cruralis*, could probably evolve fairly fast under strong selection, as has been suggested to have happened multiple times in other families within Sylvioidea (Alström et al.,

2011b, 2013b). Niche shifts, such as in *Malia grata*, have likely triggered strong morphological divergence in some other Indonesian island endemics (Fjeldså et al., 2010; Alström et al., 2015). However, it cannot be excluded that phylogenetic results suggesting unexpectedly distant relationships between morphologically similar species might be the result of stochastic processes, such as lineage sorting across multiple speciation events ('hemiplasy'; Avise and Robinson, 2008) rather than parallel evolution.

Within clade B, subclades J and K are well supported and deeply diverged. As further support of this subdivision, there are average differences in size and song between them (see below). *Bradypterus alfredi* is well supported as sister to *Locustella luscinioides* + *L. fluviatilis* within clade J. It is the only Afrotropical species in clade B.

The three species of *Robsonius* form a strongly supported sister clade to the rest of Locustellidae. This agrees with Oliveros et al. (2012), who disclosed this unexpected position of *Robsonius*, which was previously considered a “babbler” (e.g. Dickinson 2003).

4.2. Dating and intraspecific variation

Recent, broadly sampled, genome-scale studies suggest generally younger ages for sylvioid passerines than we recovered here (Prum et al. 2015; Moyle et al. 2016). However, our results largely agree with those from a multilocus phylogenetic analysis of all Himalayan passerines, which was dated using multiple fossils and biogeographic dates (Price et al., 2014), as well as with an analysis of the modern birds (Neornithes) using a multigene matrix and a large number of fossil calibrations (Claramunt and Cracraft, 2015). In particular the date inferred here for the divergence of *Robsonius* should be treated with caution.

Several polytypic species harbour deep *cytb* divergences, which are considerably deeper than between some other taxa treated as separate species. Because the morphological variation within some groups is relatively slight and because many of the species are poorly known, it seems likely that some currently recognised species are better treated as two (or more) species. Further studies of larger samples and using independent data, such as nuclear markers and vocalisations, are warranted to shed light on the taxonomic status of these taxa.

The rather widely allopatric *Bradypterus lopezi ufipae* and *B. l. mariae* are deeply diverged (2.1 mya; 95% HPD 1.3–3.0 mya). This species consists of two subspecies groups (Kennerley and Pearson 2010; del Hoyo and Collar, 2016), but *B. l. ufipae* and *B. l. mariae* belong to the same group. However, these two taxa inhabit different montane areas in central and eastern Africa. The divergence within the Madagascar endemic *Bradypterus brunneus*, which was estimated at 5.7 mya (95% HPD 4.1–7.5 mya) is exceptional, as this species is

considered monotypic (Dickinson and Christidis, 2014; del Hoyo and Collar, 2016; Gill and Donsker, 2017).

The divergence between *Megalurus palustris forbesi* from the Philippines and northern Borneo and *M. p. toklao*, which is patchily distributed across southern continental Asia, is pronounced (4.0 mya; 95% HPD 2.7–5.5 mya). Much denser sampling will be needed to evaluate the taxonomy of this complex. A deep split (2.2. mya; 95% HPD 1.4–3.1 mya) was also found between the *Megalurus punctatus caudatus* (confined to Snares Island, New Zealand) and *M. p. vealeae* (North Island, New Zealand). The former has recently been treated as a distinct species based on morphological characters (del Hoyo and Collar, 2016). Our single sample of the extinct *M. rufescens* from Chatham Island, New Zealand is deeply diverged from *M. punctatus*. These two taxa are often treated as conspecific (e.g. Dickinson and Christidis, 2014), although del Hoyo and Collar (2016) treated them as separate species based on morphological differences. However, the suggestion by del Hoyo and Collar's (2016) that *M. rufescens* might be most closely related to *Megalurulus rufus* is strongly rejected by our data.

Megalurus timoriensis and *M. macrurus* are paraphyletic with respect to each other, with deep divergences between *M. t. tweeddalei* and *M. t. crex* and between *M. m. macrurus* and *M. m. interscapularis*, respectively. The two latter taxa belong to different subspecies groups, which differ in elevational distribution and number of tail feathers (Schodde and Mason, 1999; Dickinson and Christidis 2014; del Hoyo and Collar, 2016), whereas to our knowledge the other taxa have not been suggested to be markedly different. A taxonomic revision may be warranted, but more extensive research is needed.

Locustella certhiola, *L. ochotensis* and *L. pleskei* have long been considered closely related, and have been treated variously as either conspecific or different species (review in Kennerley and Pearson, 2010). Our study supports a sister relationship and recent divergence (0.68 mya; 95% HPD 0.35–1.06 mya) between the allopatric *L. pleskei* and two of the three *L. ochotensis*, with *L. certhiola* and the third *L. ochotensis* as more deeply diverged sisters (2.9 mya; 95% HPD 2.0–4.0 mya). The *L. ochotensis* with a *L. certhiola* *cytb* haplotype was collected at Magadan, Russia (Takema Saitoh, in litt.). The same topology (except for paraphyly of *L. ochotensis/L. certhiola*) was previously found using a smaller mitochondrial dataset (Drovetski et al., 2004). A more comprehensive study by Drovetski et al. (2015) also found *L. certhiola* as sister to the two others in a mitochondrial ND2 tree, but recovered *L. pleskei* to be paraphyletic with respect to *L. ochotensis*, and also identified one phenotypic *L. certhiola* from Khabarovsk with a *L. ochotensis* ND2 haplotype. In contrast, in a species tree

based on 12 nuclear introns, the same authors recovered *L. certhiola* and *L. ochotensis* as sisters, with *L. pleskei* sister to these two, with strong support. The dating of the deepest node in their ND2 tree was considerably younger than in our study (1.6 mya; 95% HPD 1.2–2.0 mya), and the ages estimated by the nuclear introns were even younger.

4.3. Taxonomic implications

Alström et al. (2011a) found the genus-level taxonomy of Locustellidae to be strongly incongruent with the phylogeny, and proposed a major reclassification. The present analysis, which includes 13 species whose phylogenetic position has either not been tested previously or only in a narrower context, revealed further conflict between taxonomy and phylogeny. The revised taxonomy of Alström et al. (2011a) recognised only four genera: *Locustella* (comprising the traditional *Locustella* and all Asian *Bradypterus*), *Bradypterus* (restricting this genus to the African species), *Schoenicola* (only *S. brevirostris* studied) and *Megalurus* (including *Eremiornis carteri*, *Cincloramphus cruralis* and *C. mathewsi*).

Alström et al. (2011a) stressed that their proposed *Megalurus* was probably non-monophyletic, but noted that the support was based mainly on *cytb*, and that this was contradicted by other data. Therefore, the authors preliminarily retained *Megalurus* for a potentially non-monophyletic group. However, Dickinson and Christidis (2014) and del Hoyo and Collar (2016), based on the same study by Alström et al. (2011a), restricted *Megalurus* to *M. palustris* (type species of this genus). Moreover, Dickinson and Christidis (2014) and del Hoyo and Collar (2016) applied the name *Cincloramphus* to *M. cruralis*, *M. mathewsi*, *M. timoriensis* and *M. macrurus* (the two former species were previously placed in this genus). They also resurrected the genus *Poodytes* for *M. gramineus*, *M. punctatus*, *M. caudatus*, *M. rufescens*, *M. carteri* and *M. albolimbatus* (though *M. caudatus* and *M. rufescens* were not given species status by Dickinson and Christidis, 2014). *Megalurus albolimbatus* has not yet been analysed phylogenetically.

Our results call for further taxonomic revision, although due to the uncertain relationships especially with respect to *Megalurus*, none of the alternative classifications are fully satisfactory. Synonymising *Amphilais* with *Bradypterus* is straightforward, and so is moving *Bradypterus alfredi* to *Locustella*. Due to the strong support for a sister relationship between *Chaetornis striata* and *Schoenicola platyurus*, we propose synonymising the former genus with the latter (based on priority). As there is no unanimous support for a close relationship between these and *Schoenicola brevirostris*, we hesitantly propose reinstating the name *Catriscus* Cabanis, 1851 for the latter (which then becomes a monotypic genus).

Because *Elaphrornis* has no obvious close relatives, we suggest retaining this monotypic genus.

With respect to *Megalurus*, we follow Dickinson and Christidis (2014) and del Hoyo and Collar (2016) in restricting this name to *M. palustris* and applying the name *Poodytes* to clade G. We also accept the use of the name *Cincloramphus* for clade F excluding *M. grata*, but also propose to synonymise *Megalurulus* and *Buettikoferella* with *Cincloramphus*. *Malia grata* is strongly supported as belonging to clade F, and was estimated to have separated less than one million years before *C. cruralis* diverged. Accordingly, it would be appropriate to include it in *Cincloramphus* (by priority); however, to maintain taxonomic stability and also to highlight its morphological and ecological uniqueness, we prefer to retain the name *Malia*. Application of the name *Robsonius* is unproblematic.

Alström et al. (2011a) proposed placing all of the taxa in clade B in *Locustella* (except *L. alfredi*, which they did not analyse), and this was followed by Dickinson and Christidis (2014), del Hoyo and Collar (2016) and Gill and Donsker (2017). However, in order to use the genus category in a more consistent way across the two main clades of Locustellidae, we propose splitting *Locustella* into two genera. The name *Locustella* is restricted to clade J. However, no name is available for clade K. Accordingly, we here propose a new genus name for this clade:

***Helopsaltes*, new genus**

Type species: Motacilla Certhiola Pallas, 1811. Gender masculine.

Included taxa: All of the species in clade K in Figs 1 and 2, which should now be named *Helopsaltes certhiola*, *Helopsaltes ochotensis*, *Helopsaltes pleskei*, *Helopsaltes pryeri*, *Helopsaltes fasciolatus* and *Helopsaltes amnicola*. All species epithets except *fasciolatus* are invariable, and therefore must not change ending due to change of gender of the scientific name.

Diagnosis: The songs consist of short (c. 2–5 s) strophes separated by distinct pauses (c. 2–15 s; highly variable depending on level of excitement). All or most of the elements in the strophes are different from each other, or arranged in different “blocks” of similar notes. The songs of the species of *Locustella sensu stricto* are less clearly separated into strophes, and consist of very fast rattling reels or monotonous repetitions of rather simple syllables. See 3.3 and Fig. 3. No diagnostic morphological characters are known to us, but there are average differences between *Helopsaltes* and *Locustella sensu stricto* in overall size (see 3.3).

Etymology: The name means “the marsh musician”, from Greek helos (ἕλος), marshy ground, and Greek psaltes (ψάλτης), a musician playing a string instrument.

Acknowledgements

We are grateful to Joel Cracraft, Paul Sweet and Thomas J. Trombone (American Museum of Natural History); Robert Prys-Jones and Mark Adams (The Natural History Museum, Tring); Mark Robbins, University of Kansas Biodiversity Institute; Frederick H. Sheldon (LSU Museum of Natural Science Collection of Genetic Resources, Louisiana State University); Ulf Johansson and Peter Nilsson (Swedish Museum of Natural History); Christopher M. Milensky (US National Museum); Richard Prum and Kristof Zyskowski (Yale Peabody Museum of Natural History); Jan Bolding Kristensen (Zoological Museum, University of Copenhagen); Staffan Andersson; Hem Sagar Baral; Geoff Carey; Paul Leader; Ulf Ottosson; and Yoshimitsu Shigeta for providing some samples. A.C. benefitted from a Collection Study Grant from the Frank M. Chapman Memorial Fund to visit the AMNH. E.P. thanks Jean-Jérôme Cassan (Direction du développement économique et de l'environnement de la province Nord), Anne-Claire Goarant (Direction de l'environnement de la province Sud), Romain Bonifay (Municipalité de Nouméa), Julien Baudat-Franceschi (Société Calédonienne d'Ornithologie), Almudena Lorenzo and Mariane Bonzon (Parc Zoologique et Forestier de Nouméa), Péguy and Jean-Pierre Drain (Nouméa) for their assistance during fieldwork in New Caledonia. E.P. was supported by the grant ANR Biodiversité BIONEOCAL, a project led by Philippe Grandcolas (MNHN). M.J.A. and R.G.M. thank Marika Tuiwawa, Alivereti Naikatini, Sanivalati Vido, the Fiji Department of Forestry and the Biosecurity Authority of Fiji for assistance in Fiji. We are grateful to Steven Gregory, Normand David, Ib Friis, Minna Skafte Jensen, Edward Dickinson, Frank Gill, David Donsker, Alan Peterson and Les Christidis for advise on the new scientific name; to Josep del Hoyo, Elisa Badia and Lynx Edicions for allowing us to publish illustrations from Handbook of the Birds of the World; to Lars Petersson for letting us publish his photos of *Locustella pryeri*; to Dave Farrow for providing a sound recording of *Locustella major*; and to Laima Bagdonaite for submitting the sequences to GenBank. Some of the analyses were run on the CIPRES Science Gateway. P.A. was supported by the Swedish Research Council (2015-04402), Jornvall Foundation, and Mark and Mo Constantine; M.I. was supported by the Swedish Research Council (621-2014-5113); M.J.A. and R.G.M. were supported by NSF (DEB-1557051 and DEB-1557053, respectively); and J.F. acknowledges the Danish National Research Foundation for support to

the Center for Macroecology, Evolution and Climate (DNRF96). We are most grateful to Frank Rheindt and one anonymous reviewer for helpful comments on the manuscript.

References

- Alström, P., Ericson, P.G.P., Olsson, U., Sundberg, P., 2006. Phylogeny and classification of the avian superfamily Sylvioidea. *Molecular Phylogenetics and Evolution* 38, 381–397.
- Alström, P., Fregin, S., Norman, J.A., Ericson, P.G.P., Christidis, L., Olsson, U., 2011a. Multilocus analysis of a taxonomically densely sampled dataset reveal extensive non-monophyly in the avian family Locustellidae. *Molecular Phylogenetics and Evolution* 58, 513–526.
- Alström, P., Höhna, S., Gelang, M., Ericson, P.G.P., Olsson, U. 2011b. Non-monophyly and intricate morphological evolution within the avian family Cettiidae revealed by multilocus analysis of a taxonomically densely sampled dataset. *BMC Evol. Biol.* 11: 352.
- Alström, P., Olsson, U., Lei, F., 2013a. A review of the recent advances in the systematics of the avian superfamily Sylvioidea. *Chinese Birds* 4, 99–131.
- Alström, P., Barnes, K.N., Olsson, U., Barker, F.K., Bloomer, P., Khan, A.A., Qureshi, M.A., Guillaumet, A., Crochet, P.-A., Ryan, P.G., 2013b. Multilocus phylogeny of the avian family Alaudidae (larks) reveals complex morphological evolution, non-monophyletic genera and hidden species diversity. *Mol. Phylogenet. Evol.* 69, 1043–1056.
- Alström, P., Hooper, D.M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., Hung, L.M., Zhao, J., Lei, F. & Price, T.D. 2014. Discovery of a relict lineage and monotypic family of passerine birds. *Biology Letters* 10: 20131067.
- Alström, P., Jönsson, K., Fjeldså, J., Ödeen, A., Ericson, P.G.P., Irestedt, M., 2015. Dramatic niche shifts and morphological change in two insular bird species. *Roy. Soc. Open Sci.* 2: 140364.
- Avise, J. C., & Robinson, T. J. 2008. Hemiplasy: a new term in the lexicon of phylogenetics. *Systematic Biology*, 57, 503–507.
- Bock, W.J., 1994. History and nomenclature of avian family-group names. *Bull. Am. Mus. Nat. Hist.* 222, 1–283.
- Claramunt, S. & Cracraft, J. 2015. A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances* 1:e1501005.
- Collar, N., Robson, C. 2017. *Malia* (*Malia grata*). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. & de Juana, E. (eds.). *Handbook of the Birds of the World Alive*. Lynx

- Edicions, Barcelona. (retrieved from <http://www.hbw.com/node/59676> on 14 September 2017).
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods* 9, 772–772.
- del Hoyo, J., Collar, N.J., 2016. HBW and BirdLife International illustrated checklist of the birds of the world. Volume 2: Passerines. Lynx Edicions, Barcelona.
- del Hoyo, J., Elliot, A., Christies, D.A., 2006. Handbook of the Birds of the World. Vol. 11. Old World Flycatchers to Old World Warblers. Lynx Edicions, Barcelona.
- Dickinson, E.C., 2003. The Howard & Moore Complete Checklist of the birds of the World. Christopher Helm, London.
- Dickinson, E.C., Christidis, L., 2014. The Howard & Moore complete checklist of the birds of the world. 4th edition. Vol. 2. Aves Press, Eastbourne, U. K.
- Drovetski, S.V., Zink, R.M., Fadeev, I.V., Nesterov, E.V., Koblik, E.A., Red'kin, Y.A., Rohwer, S., 2004. Mitochondrial phylogeny of *Locustella* and related genera. *Journal of Avian Biology* 35, 105–110.
- Drovetski, S.V., Semenov, G., Red'kin, Y.A., Sotnikov, V.N., Fadeev, I.V., Koblik, E.A., 2015. Effects of asymmetric nuclear introgression, introgressive mitochondrial sweep, and purifying selection on phylogenetic reconstruction and divergence estimates in the Pacific clade of *Locustella* warblers. *PloS One*, 10(4), e0122590.
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol* [Internet] 4.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29, 1969–1973.
- Fjeldså, J., Irestedt, M., Ericson, P. G., Zuccon, D., 2010. The Cinnamon Ibon *Hypocryptadius cinnamomeus* is a forest canopy sparrow. *Ibis* 152, 747-760.
- Fregin, S., Haase, M., Olsson, U., Alström, P., 2012. New insights into family relationships within the avian superfamily Sylvioidea (Passeriformes) based on seven molecular markers. *BMC Evolutionary Biology* 12, 157.
- Gill, F., Donsker, D., 2017. IOC World Bird List (v 7.2). doi : 10.14344/IOC.ML.7.2.
- Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27, 570–580.
- Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.

- Irestedt, M., Ohlson, J. I., Zuccon, D., Källersjö, M. & Ericson, P. G. P., 2006. Nuclear DNA from old collections of avian study skins reveals the evolutionary history of the Old World suboscines (Aves, Passeriformes). *Zool. Scr.*, 35, 567–580.
- Johansson U.S., Fjeldså J., Bowie R.C.K., 2008. Phylogenetic relationships within Passerida (Aves: Passeriformes): a review and a new molecular phylogeny based on three nuclear intron markers. *Molecular Phylogenetics and Evolution* 48, 858–876.
- Kennerley, P.R., Pearson, D., 2010. Reed and Bush Warblers. Christopher Helm, London.
- Madge, S. 2017. Broad-tailed Grassbird (*Schoenicola platyurus*); Fan-tailed grassbird (*Schoenicola brevirostris*); and Bristled Grassbird (*Chaetornis striata*). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. & de Juana, E. (eds.). *Handbook of the Birds of the World Alive*. Lynx Edicions, Barcelona. (retrieved from <http://www.hbw.com/node/58724>, <http://www.hbw.com/node/587245>, and <http://www.hbw.com/node/58722>, respectively, on 28 July 2017).
- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 14 Nov. 2010, New Orleans, LA pp 1–8.
- Moyle, R.G., Oliveros, C.H., Andersen, M.J., Hosner, P.A., Benz, B.W., Manthey, J.D., Travers, S.L., Brown, R.M., Faircloth, B.C., 2016. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. *Nature Comm.*, 7:12709.
- Oliveros, C.H., Reddy, S., Moyle, R.G., 2012. The phylogenetic position of some Philippine “babblers” spans the muscicapoid and sylvioid bird radiations. *Molecular Phylogenetics and Evolution* 65, 799–804.
- Price, T.D., Hooper, D.M., Buchanan, C.D., Johansson, U.S., Tietze, D.T., Alström, P., Olsson, U., Ghosh-Harihar, M., Ishtiaq, F., Gupta, S.K., Martens, J.E., Harr, B., Singh, P., Mohan, D. 2014. Niche filling slows the diversification of Himalayan songbirds. *Nature* 509: 222–225.
- Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526: 569–573.
- Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6, Available from <http://beast.bio.ed.ac.uk/Tracer>
- Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.

- Schodde, R., Mason, I.J., 1999. The Directory of Australian Birds: Passerines. CSIRO Publishing, Melbourne.
- Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Watson, G.E., Traylor, M.A., Jr., Mayr, E., 1986. Family Sylviidae. In: Mayr, E., Cottrell, G.W. (Eds.), *A Check-list of birds of the world: a continuation of the work of James L. Peters*. Museum of Comparative Zoology, Cambridge, Massachusetts, pp. 3–294.
- Weir, J.T., Schluter, D., 2008. Calibrating the avian molecular clock. *Molecular Ecology* 17, 2321–2328.

Figure 1. *BEAST phylogeny of Locustellidae based on the mitochondrial *cytb*, and nuclear *myo*, *ODC*, *LDH* and *GAPDH* introns. Posterior probabilities (PP) and maximum likelihood bootstrap (MLBS) values are indicated at the nodes, in this order; * means PP 1.00/MLBS 100%. Clade labels (A–J) indicate clades discussed in the text; # indicates species not previously analysed phylogenetically; and § indicates species not analysed in previous comprehensively sampled Locustellidae phylogeny. ¹ MLBS 83% for clade with *Schoenicola brevirostris* sister to clade C. Illustrations by Ian Lewington, Brian Small, Jan Wilczur and Tim Worfolk, from del Hoyo et al. (2006), with permission from the publishers.

Figure 2. Cytochrome *b* (*cytb*) chronogram including all species and subspecies available for this study. Dates based on a 2.1%/million years molecular clock. Species for which only *cytb* is available, and which are therefore not in Figure 1, are indicated by €. Posterior probabilities (PP) are indicated at the nodes; * means PP 1.00. Clade labels (A–J) indicate clades discussed in the text. The purple line at 2 million years ago highlights the existence of several deep intraspecific divergences.

Figure 3. Sonograms of *Locustella sensu lato*. All of the species in clade K are shown, but not all of the ones in clade J, though the ones missing from clade J have songs that are reminiscent of species illustrated. The species in clade K have short complex strophes separated by pauses of varying length (length of pauses mainly depending on level of excitement); only in two of the individuals shown here are the pauses short enough to show two consecutive strophes. The species in clade J have “continuously flowing” songs, except *L. tacsanowskia*, which has well defined strophes of varying length separated by pauses of variable length. The three lowermost sonograms are shown at higher temporal resolution than the others to highlight the greater complexity and less regular structure of *L. pryeri* compared to *L. tacsanowskia* and *L. luteoventris*. The photos are a collage of a *L. pryeri* singing in song-flight, which all in clade K except *L. amnicola* and *L. fasciolata* regularly do (photos: Lars Petersson; Honshu, Japan, June 2017). See Supplementary Material S2 for details about the recordings used to create the sonograms.

Supplementary Figure S1. BEAST maximum clade credibility tree based on the concatenated mitochondrial *cytb*, and nuclear *myo*, *ODC*, *LDH* and *GAPDH* introns. Posterior probabilities (PP) and maximum likelihood bootstrap (MLBS) values are indicated at the nodes, in this order; * means PP 1.00/MLBS 100%. Values in blue are significantly

higher than in the *BEAST phylogeny (Fig. 1) (<0.95 in *BEAST, ≥ 0.95 in this tree); values in purple indicate strongly supported nodes not found in *BEAST tree; value in red indicates node with PP <0.95 in present tree, but >0.95 in *BEAST.

Supplementary Figure S2. MrBayes maximum clade credibility tree of primary Passerida groups based on the mitochondrial *cytb*, and nuclear *myo*, *ODC*, *LDH* and *GAPDH* introns. Posterior probabilities (PP) are indicated at the nodes; * means PP 1.00. This tree shows that the sister relationship between *Robsonius* and other locustellids is not due to poor sampling of outgroup species.

Supplementary Figure S3. BEAST maximum clade credibility tree based on the nuclear *myo*, *ODC*, *LDH* and *GAPDH* introns. Posterior probabilities (PP) are indicated at the nodes; * means PP 1.00.

Supplementary Figure S4. BEAST maximum clade credibility trees based on single-locus analyses of nuclear loci (except *RAG1*). Posterior probabilities (PP) are indicated at the nodes; * means PP 1.00.

Supplementary Figure S5. BEAST maximum clade credibility tree based on the nuclear *RAG1* gene (all sequences from GenBank; accession numbers after names). Posterior probabilities (PP) are indicated at the nodes; * means PP 1.00.

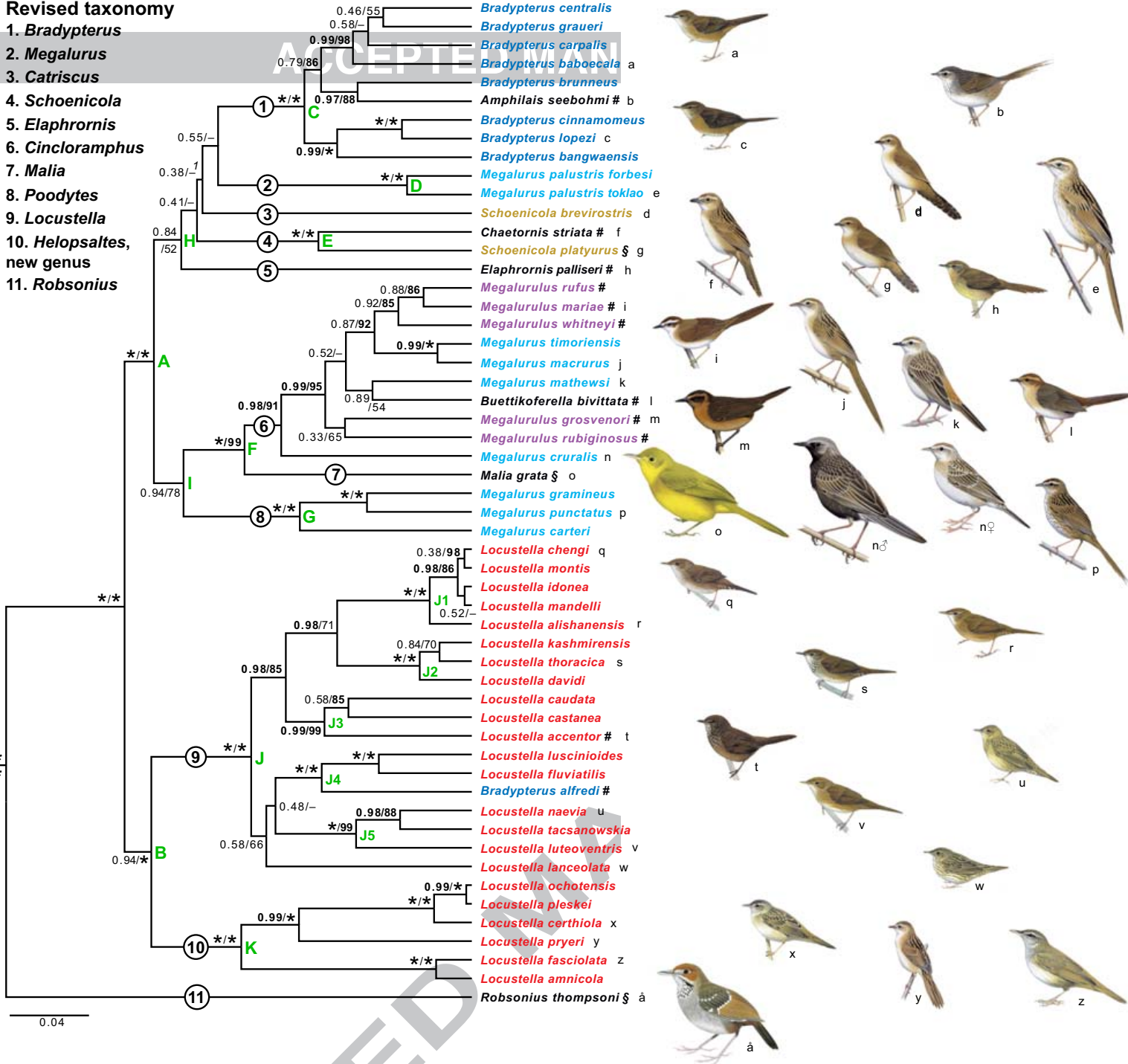
Supplementary Material S1. Xml files for analyses of Locustellidae: **S1a:** *BEAST (Fig. 1); **S1b:** BEAST all loci (except *RAG1*) concatenated (Supplementary Fig. S1); **S1c:** BEAST nuclear loci (Supplementary Fig. S3). **S1d:** Newick tree file from *BEAST analysis

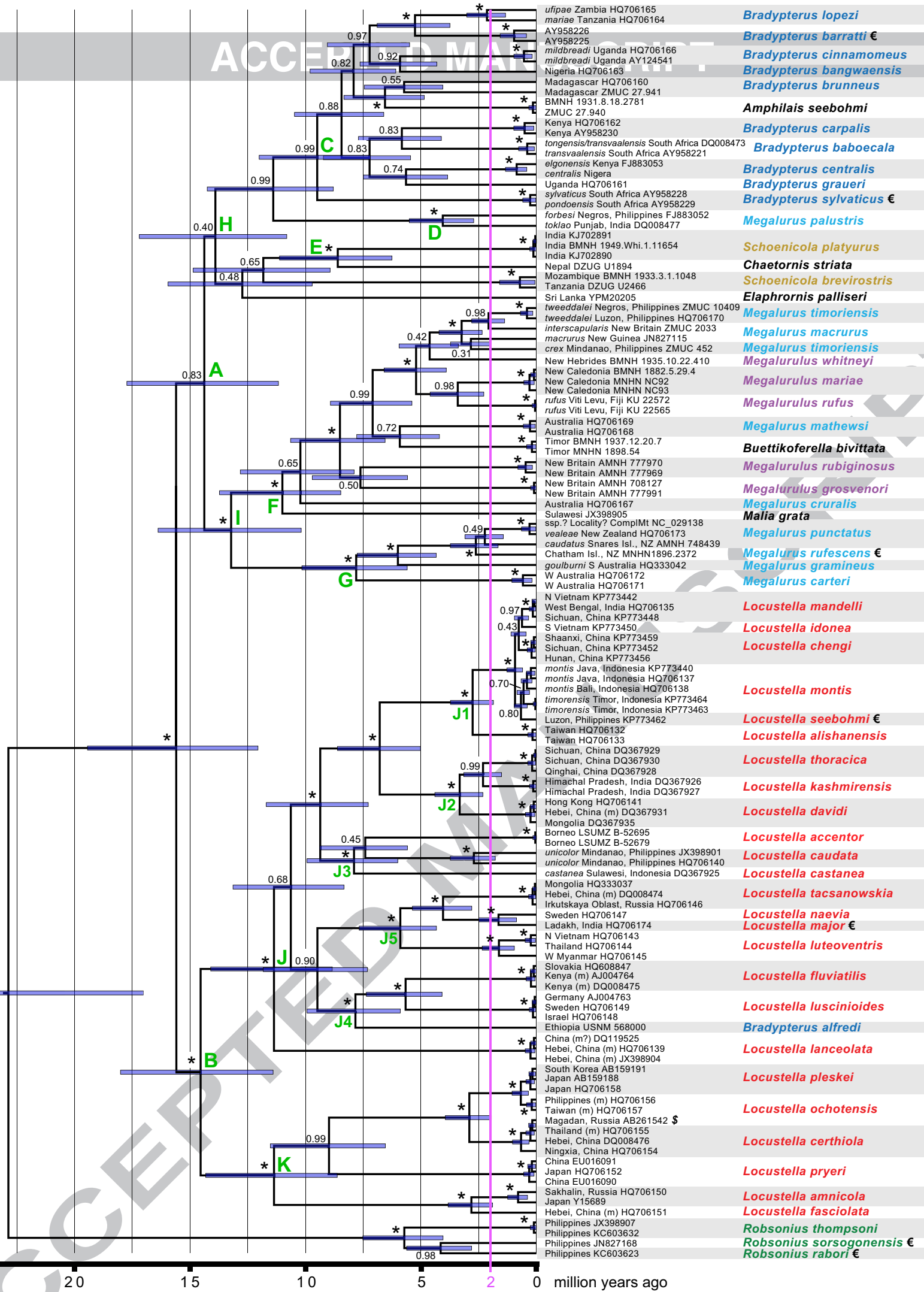
Supplementary Material S2. List of sound recordings used to create sonograms in Figure 3. XC numbers are from www.xeno-canto.org, and BL numbers from the British Library National Sound Archive. *Locustella amnicola* Kunashir Island, Russia, June 2017 (Ukolov Ilya; XC379283). *L. fasciolata* Dailing, Heilongjiang Prov., China, June 1987 (Per Alström; XC402478). *L. certhiola* Medamuji, Nei Menggu Autonomous Region, China, July 1987 (Per Alström; XC402475). *L. ochotensis* Japan (Tsuruhiko Kabaya; Songs & calls of 333 birds in Japan; 1996; Shogakkan, Tokyo). *L. pleskei* Miyake Island, Japan, June (Urban Olsson; XC402486). *L. pryeri* Zhalong, Heilongjiang Prov., China, June 1987 (Per Alström;

XC402488). *L. accentor* (Simon Harrap; BL56358). *L. caudatus* (Simon Harap; BL41240). *L. alishanensis* Taiwan, May 1999 (Per Alström; XC402442). *L. mandelli* Fansipan, Lào Cai Prov., Vietnam, May 1999 (Per Alström; XC402485). *L. kashmirensis* Manali, Himachal Pradesh, India, June 1999 (Per Alström; XC402481). *L. thoracica* Mengda, Qinghai Prov., China, June 1994 (Per Alström; XC402491). *L. davidi* Badaling, Beijing, May 1988 (Per Alström; XC402477). *L. naevia* Uppsala, Sweden June 2004 (Per Alström; XC402480). *L. major* Pakistan (Dave Farrow; unpublished). *L. tacsanowskia* Nei Menggu, China, July 2017 (Paul Holt; XC381998). *L. luteoventris* Emei Shan, Sichuan Prov., China, June 1990 (Per Alström; XC402483). *L. fluviatilis* Uppsala, Sweden June 2004 (Per Alström; XC402479). *L. lanceolata* Huzong, Heilongjiang Prov., China, June 1988 (Per Alström; XC402482). *L. luscinioides* The Netherlands, June 2017 (Joost van Bruggen; XC374537). *L. alfredi* Uganda, September 2014 (Tomas Carlberg; XC195878).

Revised taxonomy

1. *Bradypterus*
2. *Megalurus*
3. *Catriscus*
4. *Schoenicola*
5. *Elaphornis*
6. *Cincloramphus*
7. *Malia*
8. *Poodytes*
9. *Locustella*
10. *Helopsaltes*,
new genus
11. *Robsonius*

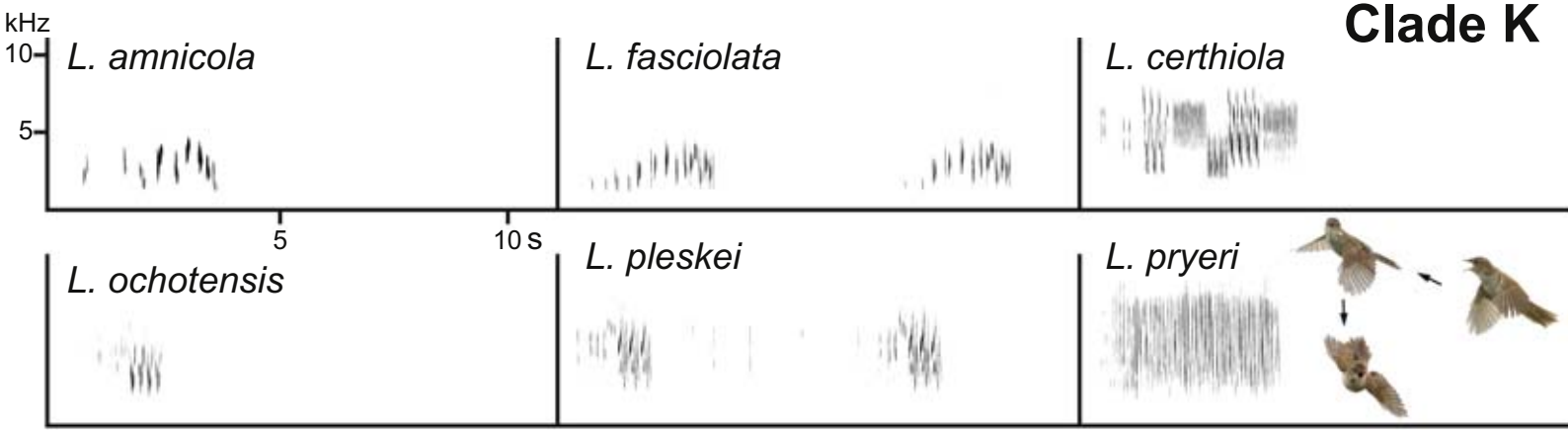




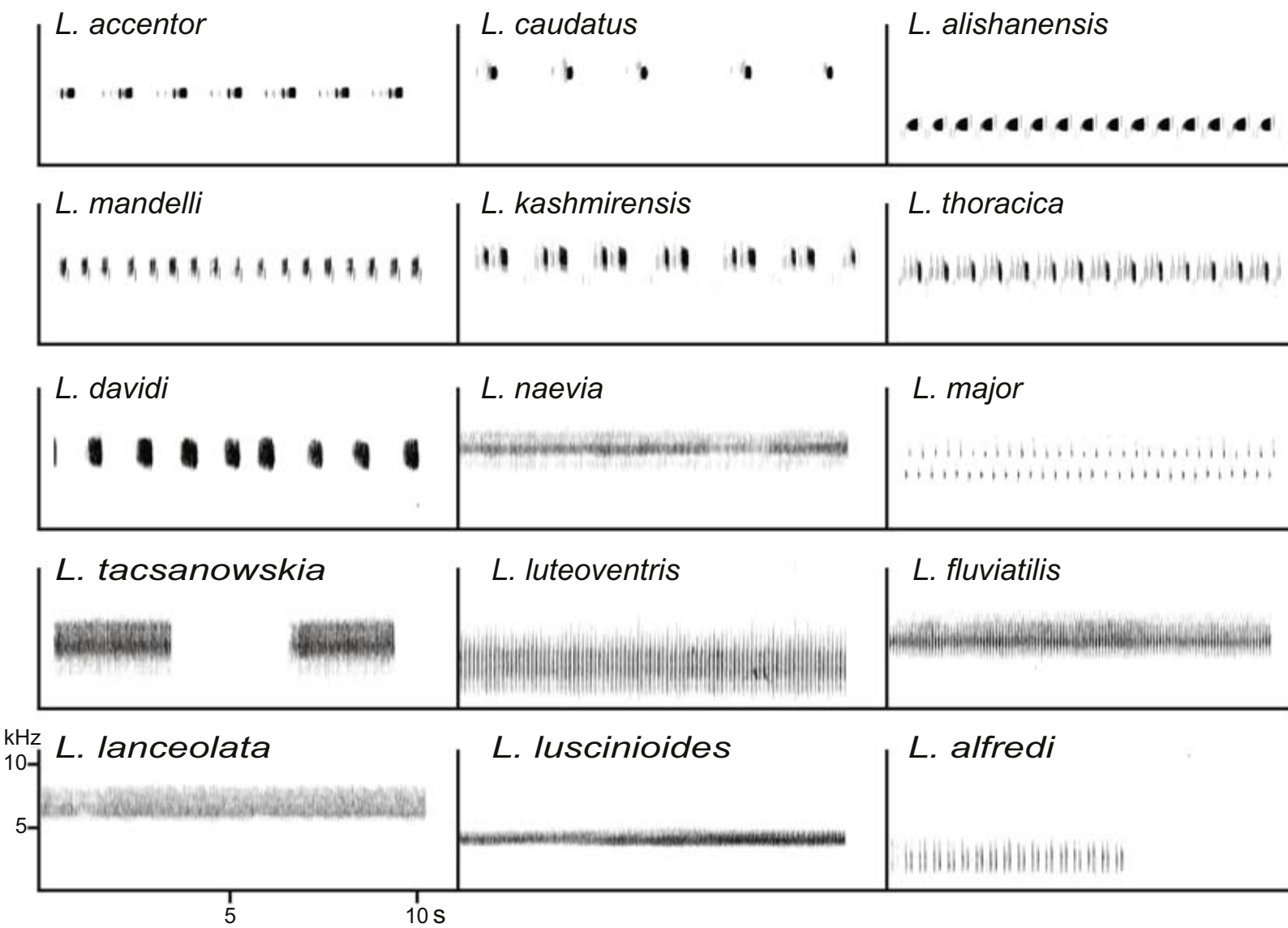
- Bradypterus lopezi*
- Bradypterus barratti* €
- Bradypterus cinnamomeus*
- Bradypterus bangwaensis*
- Bradypterus brunneus*
- Amphilais seebohmi*
- Bradypterus carpalis*
- Bradypterus baboecala*
- Bradypterus centralis*
- Bradypterus graueri*
- Bradypterus sylvaticus* €
- Megalurus palustris*
- Schoenicola platyurus*
- Chaetornis striata*
- Schoenicola brevirostris*
- Elaphornis palliseri*
- Megalurus timoriensis*
- Megalurus macrurus*
- Megalurus timoriensis*
- Megalurulus whitneyi*
- Megalurulus mariae*
- Megalurulus rufus*
- Megalurus mathewsi*
- Buettikoferella bivittata*
- Megalurulus rubiginosus*
- Megalurulus grosvenori*
- Megalurus cruralis*
- Malia grata*
- Megalurus punctatus*
- Megalurus rufescens* €
- Megalurus gramineus*
- Megalurus carteri*
- Locustella mandelli*
- Locustella idonea*
- Locustella chengi*
- Locustella montis*
- Locustella seebohmi* €
- Locustella alishanensis*
- Locustella thoracica*
- Locustella kashmirensis*
- Locustella davidi*
- Locustella accentor*
- Locustella caudata*
- Locustella castanea*
- Locustella tacsanowskia*
- Locustella naevia*
- Locustella major* €
- Locustella luteoventris*
- Locustella fluviatilis*
- Locustella luscinioides*
- Locustella lanceolata*
- Locustella pleskei*
- Locustella ochotensis*
- Locustella certhiola*
- Locustella pryeri*
- Locustella amnicola*
- Locustella fasciolata*
- Robsonius thompsoni*
- Robsonius sorsogonensis* €
- Robsonius rabori* €

20 15 10 5 2 0 million years ago

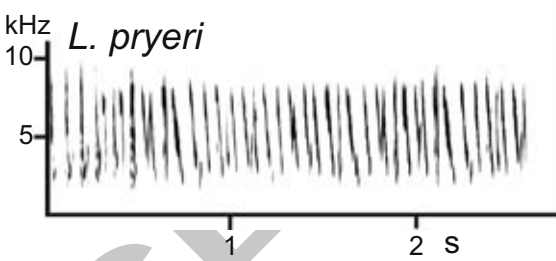
Clade K



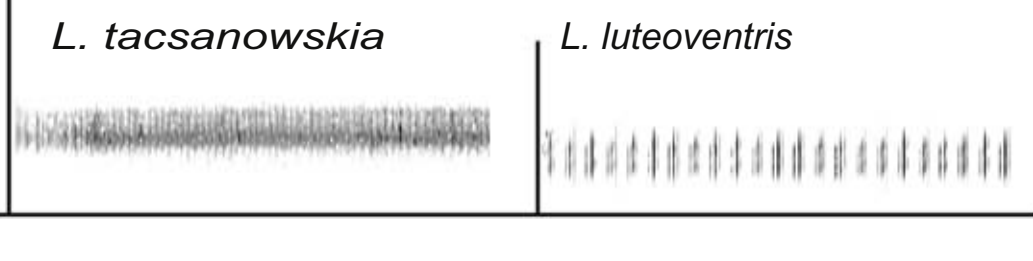
Clade J



Clade K



Clade J



Highlights

- Phylogeny of avian family Locustellidae based on multilocus data for all except 3 of the species
- Exceedingly high level of disagreement with current taxonomy due to complex morphological evolution
- Revised taxonomy proposed
- One new genus described
- Species diversity probably underestimated as indicated by multiple deep intraspecific divergences

ACCEPTED MANUSCRIPT

Revised taxonomy

1. *Bradypterus*
2. *Megalurus*
3. *Catriscus*
4. *Schoenicola*
5. *Elaphrornis*
6. *Cincloramphus*
7. *Malia*
8. *Poodytes*
9. *Locustella*
10. *Helopsaltes*,
new genus
11. *Robsonius*

