Multilocus phylogeny of the avian family Alaudidae (larks) reveals complex morphological evolution, non-monophyletic genera and hidden species diversity

Per Alström, Keith N. Barnes, Urban Olsson, F. Keith Barker, Paulette Bloomer, Aleem Ahmed Khan, Masood Ahmed Qureshi, Alban Guillaumet, Pierre-André Crochet, Peter G. Ryan

Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Chaoyang District, Beijing 100101, PR China
Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, SE-750 07 Uppsala, Sweden
Percy FitzPatrick Institute of African Ornithology, DST/NRF Centre of Excellence, University of Cape Town, Rondebosch 7700, South Africa
Systems and Biodiversity, Gothenburg University, Department of Zoology, Box 463, SE-405 30 Göteborg, Sweden
Bell Museum of Natural History, Department of Ecology, Evolution and Behavior, University of Minnesota, 1867 Upper Buford Circle, St. Paul, MN 55108, USA
Percy FitzPatrick Institute, Centre of Excellence, Department of Genetics, University of Pretoria, Hatfield 0083, South Africa
Institute of Pure & Applied Biology, Bahauddin Zakariya University, 60800 Multan, Pakistan
Department of Biology, Trent University, DNA Building, Peterborough, ON K9J 7B8, Canada
CEFE/CNRS, Campus du CNRS, 1919, Route de Mende, 34293 Montpellier, France

ARTICLE INFO

Article history:
Received 1 November 2012
Accepted 7 June 2013
Available online 21 June 2013

Keywords:
Phylogeny
Taxonomy
Morphological evolution
Nodal support

Abstract

The Alaudidae (larks) is a large family of songbirds in the superfamily Sylvioidea. Larks are cosmopolitan, although species-level diversity is by far largest in Africa, followed by Eurasia, whereas Australasia and the New World have only one species each. The present study is the first comprehensive phylogeny of the Alaudidae. It includes 83.5% of all species and representatives from all recognised genera, and was based on two mitochondrial and three nuclear loci (in total 6.4 kbp, although not all loci were available for all species). In addition, a larger sample, comprising several subspecies of some polytypic species was analysed for one of the mitochondrial loci. There was generally good agreement in trees inferred from different loci, although some strongly supported incongruences were noted. The tree based on the concatenated multilocus data was overall well resolved and well supported by the data. We stress the importance of performing single gene as well as combined data analyses, as the latter may obscure significant incongruence behind strong nodal support values. The multilocus tree revealed many unpredicted relationships, including some non-monophyletic genera (Calandrella, Mirafra, Melanocorypha, Spizocorys). The tree based on the extended mitochondrial data set revealed several unexpected deep divergences between taxa presently treated as conspecific (e.g. within Ammomanes cinctura, Ammomanes deserti, Calandrella brachyactyla, Eremophila alpestris), as well as some shallow splits between currently recognised species (e.g. Certhilauda brevirostris–C. semitorquata–C. curvirostris; Calendulauda barlowi–C. erythrochlamys; Mirafra cantillans–M. javanica). Based on our results, we propose a revised generic classification, and comment on some species limits. We also comment on the extraordinary morphological adaptability in larks, which has resulted in numerous examples of parallel evolution (e.g. in Melanocorypha mongolica and Alauda leucoptera [both usually placed in Melanocorypha]; Ammomanopsis grayi and Ammomanes cinctura/deserti [former traditionally placed in Ammomanes]; Chersophilus duponti and Certhilauda spp.; Eremopterix hova [usually placed in Mirafra] and several Mirafra spp.), as well as both highly conserved plumages (e.g. within Mirafra) and strongly divergent lineages (e.g. Eremopterix hova vs. other Eremopterix spp.; Calandrella cinerea complex vs. Eremophila spp.; Eremaluda duni vs. Chersophilus duponti; Melanocorypha mongolica and male M. yeltoniensis vs. other Melanocorypha spp. and female M. yeltoniensis). Sexual plumage dimorphism has evolved multiple times. Few groups of birds show the same level of disagreement between taxonomy based on morphology and phylogenetic relationships as inferred from DNA sequences.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The family Alaudidae, larks, comprises 97 species in 21 genera (Gill and Donsker, 2012; Spottiswoode et al., 2013), including the
Eurasian Skylark *Alauda arvensis* (“the lark”), which is familiar to many Europeans because of its widespread occurrence in agricultural land, local abundance, and beautiful song. Many other species of larks are well known for similar reasons. Larks are found on six continents, but the family's distribution and diversity is highly skewed. In terms of current distribution and diversity, the Alaudidae is primarily an African and secondarily a Eurasian family. Seventy-eight species occur in Africa, with 60 endemic to sub-Saharan Africa. Eurasia has 37 species, with one, *Mirafra javanica*, extending its range to Australia, as the only representative of this family on that continent (de Juana et al., 2004; Gill and Donsker, 2012). A single widespread species, the Horned Lark *Eremophila alpestris*, is native to the New World as well as much of the Paleartic. All 21 genera are represented in Africa, with 13 in Eurasia and one each in Australasia and the New World (de Juana et al., 2004; Gill and Donsker, 2012). In Africa, lark species richness is greatest in semi-arid and arid regions (Dean and Hockey, 1989). There are two primary centres of endemism, one in the north-east arid zone (Kenya, Ethiopia and Somalia), where 23 of the 34 species are endemic or near-endemic, and another one in the south-west arid zone (South Africa, Namibia and Botswana), where 26 of the 31 species are endemic or near-endemic (de Juana et al., 2004).

Most lark species share a similar plumage pattern: brownish or greyish above and paler below, with variously distinct darker streaking on the upperparts and breast. This pattern provides camouflage in the open, grassy or arid habitats where larks occur, and several authors have noted a positive correlation between the coloration of the upperparts of a species and the colour of the soil on which it lives (Bannerman, 1927; Guillamet et al., 2008; Kleinschmidt, 1907, 1912; Meinertzhagen, 1951; Niethammer, 1940; Vaurie, 1951). In most species, there is no sexual dimorphism in plumage, although males average larger than females. However, in *Melanocorypha yeltoniensis* and the *Eremopterix* species, male and female plumages are strongly different (and in the former, males average 13–14% heavier than females; Cramp, 1988; de Juana et al., 2004). In contrast to their cryptic plumages, most species have well developed songs, and some species, e.g. *Alauda arvensis*, are renowned songsters. Most species also have elaborate song flights. Presumably in association with diet (e.g., many species consume seeds in addition to arthropod prey), bill morphology varies considerably among species, and in some species, also between the sexes (e.g. *Alauda raza* and the long-billed lark complex; Burton, 1971; Cramp, 1988; Donald et al., 2007; Ryan and Bloomer, 1999).

Morphologically, the family Alaudidae constitutes a well-defined group, whose members share unique features of the syrinx (Ames, 1971) and tarsus (Rand, 1959). As a result, the limits of the family are not disputed, but the relationships between the larks and other taxa have long been uncertain. Linear classifications have generally placed them at the beginning of the oscine passerines (e.g. del Hoyo et al., 2004; Peters, 1960), whereas based on sequence data from a small number of mostly African species (Tieleman et al., 2003). The present study is the first comprehensive phylogeny of the Alaudidae (although part of the data for the African and some of the Western Paleartic species have been analysed in an unpublished PhD thesis; Barnes, 2007). It is based on two mitochondrial and three nuclear loci (in total 6.4 kbp, although not all loci are available for all species), and includes representatives from all recognised genera and 83.5% of all species. We also analyse one mitochondrial locus for a larger sample, comprising multiple individuals and several subspecies of some polytypic species. These data provide the basis for a major reassessment of lark relationships and taxonomy, as well as the foundation for comments on the morphological evolution in this bird family.

2. Material and methods

2.1. Study group and sampling

Taxonomy follows Gill and Donsker (2012), except with respect to *Heteromira* (*sidamoensis*, which we treat as conspecific with *H. archeri* based on Spottiswoode et al. (2013). We included 81 of the 97 species, representing all 21 genera. Eight African *Mirafra* spp., three African *Calandrella* spp. and the African *Alaemon hamertoni*, *Eremopterix leucotis* and *Spizocorys obibiensis*, as well as the Asian *Ammomanes phoenicura* and *Galerida deva* were missing.

Fresh tissue and blood samples, as well as a few feather samples, were collected by people with extensive field experience with these larks (mainly the authors of this study). Liver, heart and pectoral muscle were dissected for tissue samples, and stored in 20% dimethylsulphoxide (DMSO) and saturated salt (NaCl) (Amos and Hoezel, 1991) or ethanol. Blood samples were mixed immediately in a blood storage buffer (0.1 M Tris–HCl, 0.04 M EDTA, Na2, or 1.0 M NaCl, 0.5% SDS). Samples were refrigerated as soon as possible. Feathers were kept at −20 °C. Voucher specimens were deposited in various institutions (Appendix 1). For blood and feather samples, photographs were taken of some birds (Appendices 1 and 2). Unfortunately, a hard drive with photos of a large proportion of the species collected in Africa by KB, for which no specimens are available, has been lost.

2.2. DNA extraction and sequencing

Lab work was done mainly at the University of Pretoria (UP), University of Gothenburg (GU) and University of Minnesota (UMN). At UP DNA extractions followed standard procedures of chemical digestion, phenol/chloroform clean-up and ethanol pre-
cipitation (Sambrook et al., 1989). DNA was eluted in Sambax® (Adcock Ingram) water and stored at −20°C. At GU and UMN, DNA was extracted using QIA Quick DNEasy Kit (Qiagen, Inc.) according to the manufacturer’s instruction, but with 30 µl 0.1% DTT added to the initial incubation step of the extraction of feathers.

We sequenced five loci: the main part of the mitochondrial cytochrome b gene and part of the flanking tRNA-Thr (together referred to as cyt b); the mitochondrial 16S rRNA; the nuclear ornithine decarboxylase (ODC) exon 6 (partial), intron 6, exon 7, intron 7 and exon 8 (partial); the entire nuclear myoglobin (myo) intron 2, and the nuclear recombination activating gene, parts 1 and 2 (RAG). At GU, amplification and sequencing of cyt b followed the protocols described in Olsson et al. (2005). At UP, cyt b was amplified and sequenced using primers L14841 and H15696 and L15408 and H15915 (Edwards et al., 1991; Kocher et al., 1989; Barker et al., 2008). The amplified and sequenced data were described in Irestedt et al. (2006) were followed, with specially designed primers (Supplementary Table 1).

For ODC, and Barker et al. (2004) for RAG. Amplification and sequencing followed the protocols described for cyt, except for the modification of the primer annealing temperature (58 °C, 30 s). Amplification and sequencing followed the protocols described in Olsson et al. (2005) for myo, Allen and Omland (2003) for ODC, and Barker et al. (2004) for RAG.

DNA was also extracted from toepad samples of two Pinacorcyos species, for which no fresh DNA was available. For extraction, PCR-amplification, and sequencing procedures for these, the procedures described in Irestedt et al. (2006) were followed, with specially designed primers (Supplementary Table 1).

2.3. Phylogenetic analyses

We followed a hierarchical sampling scheme prioritizing mtDNA sampling for all species, and nuclear loci for a subset of samples, representing major lineages of larks (e.g., Wiens et al., 2005). The following sequence data were included in the analyses: cyt b for all species; 16S for nearly all African species and a few Eurasian species; and between one to three nuclear loci for most species. In addition, we analysed 142 cyt b haplotypes, including some sequences from GenBank, comprising several sub-species of polytypic species. For one species, only cyt b was available, and for 20 species, only cyt b and 16S were available. See Appendix 1 and Fig. 1 for details regarding coverage of loci across the taxa. All new sequences have been deposited in GenBank (Appendix 1).

Sequences were aligned using Muscle (Edgar, 2004) in Seaview 4.3.4 (Gouy, 2012; Gouy et al., 2010); some manual adjustment was done for the non-coding sequences. For the nuclear loci, heterozygous sites were coded as ambiguous. Trees were estimated by Bayesian inference (BI) using MrBayes 3.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) as follows: (1) All loci were analysed separately (single-locus analyses, SLAs). (2) Sequences were also concatenated, partitioned by locus (in total 5 partitions), using rate multipliers to allow different rates for different partitions (Nylander et al., 2004; Ronquist and Huelsenbeck, 2003). We also ran analyses where, in addition to the five locus-specific partitions, the coding sequences were partitioned by codon (in total 9 partitions). (3) All analyses were run under the best-fit models according to the Bayesian Information Criterion (BIC), calculated in jModeltest 0.1.1 (Posada, 2008a,b), as well as (4) using the “mixed” command to sample across the GTR model space in the Bayesian MCC (Huelsenbeck et al., 2004), and assuming rate variation across sites according to a discrete gamma distribution with four rate categories (Γ; Yang, 1994) and an estimated proportion of invariant sites (I; Gu et al., 1995). For cyt b, 16S and RAG, the model selected by the BIC was the general time-reversible (GTR) model (Lanave et al., 1984; Rodríguez et al., 1990; Tavaré, 1986) + I + Γ. For myo and ODC, the HKY model (Hasegawa et al., 1985) + Γ was chosen by the BIC. Ambiguous base pairs and indels were treated as missing data, but indels were plotted on the trees a posteriori. Pinarthus biarmicus and Prinia bairdii were chosen as outgroups based on the results of Alström et al. (2006), Johansson et al. (2008) and Fregin et al. (2012), except in the SLA of 16S, for which Cisticola brachyptera, Prinia bairdii, Acrocephalus arundinaceus and Aegithalos concinnus were used as outgroups (three latter downloaded from GenBank), as no 16S sequences were available for P. biarmicus. Default priors in MrBayes were used. Four Metropolis-coupled MCMC chains with incremental heating temperature 0.1 or 0.05 were run for 5–40 × 10^6 generations and sampled every 1000 generations. Convergence to the stationary distribution of the single chains was inspected in Tracer 1.5.0 (Rambaut and Drummond, 2009) using a minimum threshold for the effective sample size. The joint likelihood and other parameter values reported large effective sample sizes (>1000). Good mixing of the MCMC and reproducibility was established by multiple runs from independent starting points. Topological convergence was examined by eye and by the average standard deviation of split frequencies (<0.005). The first 25% of generations were discarded as “burn-in”, well after stationarity of chain likelihood values had been established, and the posterior probabilities were calculated from the remaining samples (pooled from the two simultaneous runs).

The cyt b data set with multiple subspecies was analysed in BEAST version 1.7.4 (Drummond and Rambaut, 2007, 2012). XML files for the BEAST analyses were generated in BEAUti version 1.7.4 (Rambaut et al., 2012). Analyses were run under the GTR + I model (cf. Weir and Schluter, 2008), using a “birth–death incomplete sampling” prior, and (a) a fixed clock rate of 2.13%/MY (Weir and Schluter, 2008) or (b) an uncorrelated lognormal relaxed clock (Drummond et al., 2006) with the same mean rate. Other priors were used with default values. For these analyses, 30 × 10^6 generations were run, sampled every 1000 generations. Every analysis was run twice. The MCMC output was analysed in Tracer version 1.5.0 (Rambaut and Drummond, 2009) to evaluate whether valid estimates of the posterior distribution of the parameters had been obtained. The first 25% of the generations were discarded as “burn-in”, well after stationarity of chain likelihood values had been established. Trees were summarized using TreeAnnotator version 1.7.4 (Rambaut et al., 2012), choosing “Maximum clade credibility tree” and “Mean heights”, and displayed in FigTree version 1.3.1 Rambaut (2009). The concatenated data were estimated by maximum likelihood bootstrapping (MLBS) and parsimony bootstrapping (PBS). MLBS (1000 replicates) was conducted with RAxML-HPC2 version 7.3.2 (Stamatakis, 2006; Stamatakis et al., 2008) on the Cipres portal (Miller et al., 2010). The data were partitioned by locus, and as per default GTRCAT was used for the bootstrapping phase, and GTRgamma for the final tree inference. PBS was performed in PAUP* version 4.0b10 (Swofford, 2002) on the complete dataset, using a heuristic search strategy, 1000 replicates, starting trees obtained by stepwise addition (random addition sequence, 10 replicates), TBR branch swapping, and MulTrees option not in effect (only one tree saved per replicate).

2.4. Summary of abbreviations

BI – Bayesian inference; cyt b – cytochrome b gene and part of the flanking tRNA-Thr; MLBS – maximum likelihood bootstrapping;
Fig. 1. Majority rule (50%) consensus tree of Alaudidae based on concatenated nuclear ODC, myoglobin and RAG1 + 2 and mitochondrial cytochrome b (cyt b) and 16S sequences, inferred by Bayesian inference, analysed in five partitions (one per locus; all mixed + C + I). Colours of names indicate position incongruent with current taxonomy (Gill and Donsker, 2012). Labelled bars denote clades discussed in text. Pie charts indicate posterior probabilities (PP) in single-locus analyses (see explanation in upper left corner). Support values are indicated at the nodes, in the order PP/maximum likelihood bootstrap (MLBS)/parsimony bootstrap (PBS); an asterisk represents support 1.0/100%. Red values indicate strongly supported clades that are considered uncertain despite high statistical support (see text). Coloured boxes to the right indicate sequences available for each species (see explanation in upper left corner).1 Strong conflict means PP > 0.95 for alternative relationship than the one in this figure.2 Strongly contradicted in analysis of extended cyt b dataset (cyt bE; Fig. 2). 3PP > 0.95 in cyt bE. M. yeltoniensis + M. calandra are supported as sisters with PP 1.00 in SLA of RAG, whereas M. mongolica is outside Melanocorypha clade (not strongly supported). 4MLBS and PBS infers A. arvensis + A. gulgula with 72% and 67%, respectively. 5PP 0.66 in cyt bE. 7PP 0.81 in cyt bE. Encircled numbers at nodes represent indels: (1) + 1 bp myo; (2) + 1 bp myo; (3) + 5 bp ODC; (4) + 1 bp 16S, myo; (9) − 4 bp myo; (10) − 1 bp myo, ODC, + 4 bp myo. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

myo – myoglobin intron 2; ODC – ornithine decarboxylase (mainly) introns 6–7; PBS – parsimony bootstrapping; PP – posterior probability; RAG – recombination activating gene, parts 1 and 2; SLA – single-locus analysis.
3. Results

3.1. Sequence characteristics

We obtained a contiguous ≤1002 bp of cyt b, ≤1016 bp of 16S, ≤729 bp of myo, ≤712 bp of ODC and ≤2878 bp of RAG. No unexpected stop codons or indels that would indicate the presence of nuclear pseudogenes were found in the coding sequences, although two three-bp and one six-bp indels were found in the aligned RAG sequences. The aligned cyt b sequences comprised 1002 characters, of which 439 (43.8%) were parsimony informative; 165 1016 characters, 146 (14.4%) parsimony informative; 761 characters, 115 (15.1%) parsimony informative; 746 characters, 148 (19.8%) parsimony informative; and RAG 2878 characters, 218 (7.6%) parsimony informative. The total data set comprised 6403 characters, of which 1066 (16.6%) were parsimony informative.

3.2. Concatenated multilocus analyses

The tree based on the concatenated multilocus data (Fig. 1) was overall well resolved and well supported by the data. There were three strongly supported primary clades (A–C), of which A and B were inferred to be sisters with high support. Clade A contained the monotypic genus Calandrella ("short-toed larks"), Melanocorypha, Eremophilis ("horned larks"), Galerida ("crested larks"), Alauda ("skylarks"), Lullula (Woodlark), Chersophilus (Dupont's Lark) and Eremalauda (Dunn's Lark; Sahara/Arabia), as well as the Afrotropical Spizocorys and Pseudalaemon (Short-tailed Lark). Clade B included the Afrotropical–Oriental Mirafla (bushlarks) and Afrotropical Calendulauda and heteromirafra. Clade C comprised the Afrotropical Certhidea ("long-billed larks"), Cercomorphus (Spike-heeled Lark), Pinarocorys ("thrust-like larks") and Ammomanopsis (Gray's Lark), the single Malagasy Mirafla (Madarascar Lark), the Palearctic–Afrotropical–Oriental Eremopterix ("sparrow-larks"), Ammomanes ("desert larks") and Alaemon ("hoopoe-larks"), and the Palearctic Ramphocoris (Thick-billed Lark).

Clade A could be subdivided into the strongly supported A1 and A2 (although A1 was contradicted by ODC; see Section 3.2). Clade A1 contained Calandrella, Melanocorypha, Eremophilis and the two monotypic genera Eremalauda and Chersophilus. The genus Calandrella was non-monophyletic, as some of its members (A1a) formed the sister clade to Eremalauda/Chersophilus (A1b), whereas the other members of this genus (A1d) were most closely related to Eremophilis (A1c). Also the genus Melanocorypha was non-monophyletic, as five of its species were in clade A1c, whereas the sixth species (M. leucophaea) was in A2b. Clade A2 comprised, in addition to the single Melanocorypha species, the genera Galerida (A2a), Alauda (A2b) and Spizocorys, as well as the two monotypic genera Pseudalaemon and Lullula (A2c); Pseudalaemon was nested among the Spizocorys species, whereas Lullula was sister to the others in clade A2c. The Palearctic A2a and A2b were sisters, separated from the Afrotropical (except Lullula) A2c.

Clade B could be separated into B1 and B2, both of which were strongly supported by the data. B1 comprised the monophylic genera Pinarocorys (Africa and Asia) except the Malagasy M. hova and, as sister to these, the genus Heteromirafra. The Mirafla species formed four well supported clades (B1a–B1d). The rather poorly resolved clade B2 only contained the genus Calendulauda. Within this clade, clades B2a and B2b were well supported.

Clade C could be subdivided into the well supported clades C1 and C2. Clade C1 contained Eremopterix and Mirafla hova (C1a); the genus Eremopterix was non-monophyletic, although this was poorly supported, with conflicting reconstructions in different SLAs (see Section 3.3). Clade C1b comprised Ammomanes, Pinarocorys and the monotypic Ramphocoris. In clade C2, Certhidea (C2a), Chersomanes (C2b) and the monotypic genus Ammomanopsis formed a clade that was in effect trichotomous, with Alaemon alaudipes strongly supported as sister to these taxa.

3.3. Single-locus analyses

The trees based on single-locus analyses (SLAs) of single sequences per species varied in resolution: 77.8% of the nodes in the ingroup were bifurcating in the cyt b tree, 78% in the 16S tree, 72.6% in the ODC tree, 56.8% in the myo tree and 94.6% in the RAG tree (Supplementary Fig. 1; see also Fig. 1, where SLAs are shown in pie charts). Only the cyt b tree contained the complete set of species. There were a number of topological conflicts, which received >0.95 posterior probability (PP) in different SLAs (indicated by red pie wedges in Fig. 1): (1) Calandrella ruficollis and C. alaudipes were sisters in the cyt b and myo (PP 0.97) and myo (PP 1.00) trees, whereas C. ruficollis and C. cheleensis were sisters according to ODC (PP 1.00) (data incomplete for other loci); (2) RAG supported clade A1 (PP 1.00), whereas ODC supported a clade comprising A1d, A1e and A2 (PP 0.97) (other loci unresolved; however, the extended cyt b dataset inferred a clade with A1a–A1e + A2 with PP 0.99; cf. Fig. 2); (3) cyt b, myo and RAG supported a sister relationship between clades A and B (PP 0.75, 0.93 and 0.97, respectively; cyt b was raised to 1.00 in the extended dataset, cf. Fig. 2), and myo and RAG supported clade C (PP 0.91 and 1.00, respectively), whereas clade C1 was part of the A + B clade according to ODC (PP 0.98); (4) Mirafla passerina formed a clade with M. cheniana, M. cantillans and M. javanica in the 16S tree (PP 0.95), whereas it was sister to M. williamsi in the ODC tree (PP 1.00) (cyt b unresolved, myo and RAG incomplete); (5) clades B1a–B1c formed a clade according to 16S, myo and ODC (PP 0.96, 1.00 and 0.98, respectively; cyt b unresolved), whereas RAG supported M. apiata from clade B1d as sister to clade B1c (PP 1.00); (6) Calendulauda barlowi, C. erythrochlamys and C. burra formed a clade according to cyt b (PP 0.97), whereas 16S supported C. barlowi, C. erythrochlamys and C. albescens as a clade (PP 0.99) (data incomplete for other loci); (7) Mirafla hova was part of a clade containing all Eremopterix species except E. australis in the cyt b tree (PP 0.99), whereas E. australis, not M. hova, was sister to the other Eremopterix species in the 16S (PP 0.99) and RAG trees (PP 0.97; only E. leucopareia included of “other” Eremopterix), and according to ODC, M. hova and E. australis were more closely related to clade C1b (PP 0.96) than to the two other Eremopterix species included (E. leucopareia, E. nigriceps).

3.4. Indels

Several clades were supported by apparently synapomorphic indels in the alignments of 16S, myo and ODC (Fig. 1). All of these indels supported clades that received high PPs. In addition, the sister relationship between Mirafla hova and Eremopterix australis inferred by ODC but not by any other SLA or analysis of concatenated sequences (see Section 3.2), was supported by three unique indels: a 4 bp deletion in the myo alignment and two 2 bp insertions in the ODC alignment.

3.5. Extended cytochrome b dataset

The dated tree containing multiple cyt b sequences for many species, including several subspecies (Fig. 2), basically agreed with the cyt b tree with single individuals of each species. Some nodes with PP < 0.95 in the latter tree received PPs > 0.95 in the extended dataset (indicated by footnote numbers in Fig. 1). The
youngest split between widely sympatric, reproductively isolated sister species (the Asian *Melanocorypha maxima* and *M. mongolica*; de Juana et al., 2004) was dated to 3.0 million years ago (MYA) (95% HPD 2.0–4.1 MYA) (indicated by red line in Fig. 2). The most recent split between marginally sympatric, reproductively isolated species (*Galerida cristata* and *G. macrorhyncha*; Guillaumet et al., 2005, 2006, 2008) was estimated to 1.9 MYA (95% HPD 1.3–2.7 MYA; indicated by orange line in Fig. 2). A few allo-/parapatric taxa treated as separate species were inferred to be considerably younger than this (youngest pair, *Certhilauda brevirostris*–*C. semitragula*).
orfquata, dated to 0.8 MYA, 95% HPD 0.4–1.3 MYA; indicated by purple line in Fig. 2). In contrast, several allo-/parapatric taxa treated as conspecific (in one case even consubspecific) were inferred to have diverged much longer ago. The deepest split, between Calandrella b. brachydactyla/C. b. rubiginosa and C. b. dukhuensis, which were not even inferred to be sisters, was dated to 6.0 MYA (95% HPD 4.6–7.5 MYA; indicated by blue line in Fig. 2).

4. Discussion

4.1. Phylogeny

4.1.1. Large-scale topology

This is the first comprehensive molecular study of relationships in the family Alaudidae. The only previously published study (Tielman et al., 2003) was based on cyt b and 16S for 22 species. However, nearly all of the cyt b and all of the 16S sequences of the African and some of the Western Palearctic species presented in this study, as well as some RAM sequences for exemplars from major lineages, were analysed in an unpublished PhD thesis (Barnes, 2007). The findings of this thesis formed the basis of several novel generic allocations presented in handbooks over the last decade (de Juana et al., 2004; Hockey et al., 2005). The phylogenetic hypothesis in Fig. 1 is mostly well resolved and well supported by the data, although some clades (notably A2c, B1a, B2 and C1a) include several polytomies or poorly supported nodes. The primary clades A–C, as well as the sister relationship between A and B, are strongly supported.

4.1.2. Clade A

Although clade A1 is strongly supported by the concatenated data (PP 1.00, MLBS 93%, PBS 89%), it is only recovered in one SLA (RAM) and is strongly contradicted by the SLA of ODC and by the analysis of the extended cyt b dataset. Moreover, the topologies of the ODC and cyt b trees differ from each other, resulting in three strongly supported incongruent topologies. Accordingly, clade A1 should be considered highly uncertain despite the high statistical support. This underscores the importance of critical evaluation of results, rather than just accepting high support at face value. It is possible that a species tree approach could have reconciled the incongruence among the gene trees, if it was caused by hemiplasy (reviewed by Avise and Robinson, 2008; Degnan and Rosenberg, 2009; Edwards, 2009; Liu et al., 2009). However, our data are not suitable for species tree analysis, as most species are just represented by single samples, and not all loci are available for all species. In contrast to clade A1, clade A2 is recovered with high confidence.

Within clade A1, the unexpected sister relationships between the two monotypic genera Chersophilus and Eremlaula (A1b) and between this clade and the Calandrella rufescens–cheleensis–rayatil–athensis complex (A1a) are well supported by the data. The strongly supported sister relationship between the Calandrella cinerea–brachydactyla–acrostiroclus complex (A1d) and Eremlaula (A1e) is equally surprising. All of these relationships are recovered in SLAs of two unlinked loci and are not contradicted by any other SLAs, and the A1d + A1e clade also receives support from an indel in SLAs of two unlinked loci and are not contradicted by any other results, rather than just accepting high support at face value. It is possible that a species tree approach could have reconciled the incongruence among the gene trees, if it was caused by hemiplasy. Also, S. fringillaris has been placed in a monotypic genus, Botha (Wolters, 1979; Pützold, 2003). Meinertzhagen (1951) placed S. fringillaris, S. conirostris, S. corinorhyncus, and S. personata in Calandrella. Our data refute a close relationship between any of the Spizocorys species and Calandrella or Eremlaula.

A close relationship between Galerida, Alauda and Melanocorypha leucopha (clade A2a + b) is supported by all loci. Melanocorypha leucopha is firmly nested in this clade, and hence far removed from the other Melanocorypha (A1c). The sister relationship with Alauda receives high PP and moderate bootstrap support, although this is only supported by ODC in the SLAs. This is further supported by a closer resemblance to Alauda than to Melanocorypha or Galerida in morphology, vocalisations, behaviour and ecology (see Section 4.4). Galerida magnirostris and G. modesta have been placed in the monotypic genera Calendula (Pützold, 2003; Wolters, 1979) and Helicorays (Wolters, 1979), respectively.

The generic affinity of the Raso Island (Cape Verde) endemic Alauda raza has long been unsettled. This species has been placed in Spizocorys (Boyd Alexander, 1888), Calandrella (Meinertzhagen, 1951; Pützold, 2003; Peters, 1960; Vaurie, 1959), Alauda (Wolters, 1979), or Alauda (Dean et al., 1992; Dickinson, 2003; de Juana et al., 2004; Gill and Donker, 2012; Hall, 1963), and Voous (1977) argued that its affinities are with African larks (e.g. Pseudalaemon). Hazevoet (1989, 1995) supported the placement in Alauda based on similarities with that genus in song, calls and displays (including song-flight). The molecular data corroborate this. However, our data are inconclusive with respect to the relationships among the three species of Alauda, although MLBS (72%) and PBS (67%) suggest that A. arvensis and A. gulgula are sisters.

The clade containing the five Spizocorys species (A2c) and the Short-tailed Lark Pseudalaemon fremantli is strongly supported, although for half of these only cyt b and 16S are available. The latter is usually placed in a monotypic genus (Dean et al., 1992; de Juana et al., 2004; Dickinson, 2003; Gill and Donker, 2012; Pützold, 2003; Peters, 1960; Wolters, 1979), whereas S. starki has variously been placed in Calandrella (Meinertzhagen, 1951; Peters, 1960; Wolters, 1979) or Eremlaula (Dean, 1989; Dean et al., 1992; Dickinson, 2003). The placement of S. starki in Spizocorys by de Juana et al. (2004) and Hockey et al. (2005) was based on unpublished mitochondrial DNA data from Barnes (2007). Also S. fringillaris has been placed in a monotypic genus, Botha (Wolters, 1979; Pützold, 2003). Meinertzhagen (1951) placed S. fringillaris, S. conirostris, S. corinorhynus, and S. personata in Calandrella. Our data refute a close relationship between any of the Spizocorys species and Calendula or Eremlaula.

The sister relationship between the sub-Saharan Spizocorys/ Pseudalaemon and Western Palearctic monotypic genus Lullula is well supported. Previous authors have debated whether Lullula should be recognised or synonymised with Alauda (de Juana et al., 2004; Harrison, 1966; Meinertzhagen, 1951), and Tielman et al. (2003) inferred a sister relationship between Lullula and Alau da arvensis based on cyt b and 16S. However, the present study refutes a close relationship between Lullula and Alauda.

4.1.3. Clade B

The sister relationship between the Mirafrax/Heteromirafrax clade (B1) and the Calendulauda clade (B2) is strongly supported (albeit only inferred by two SLAs, one with PP < 0.95, one with PP > 0.95), as is the sister relationship between Mirafrax and Heteromirafrax. The close relationship between the two major clades was partly unexpected, although three of the Calendulauda species have previously been placed in Mirafrax (see below). A close affinity between Mirafrax and Heteromirafrax has formerly been assumed (Dean et al., 1992), and the latter genus has been synonymized with the former (Pützold, 2003).

Within Mirafrax, the four clades B1a–B1d are recovered with a high degree of confidence. The close relationship between the five
Asian species in clade B1a is unsurprising, as they are all morphologically very similar, and four of them have been treated as conspecific (see Section 4.3). However, the relationships among these are mostly unsupported, and only cytB provides slight resolution in the SLAs. Clade B1b comprises a mix of African and Asian/Australasian taxa, including the extremely widespread M. cantillans and M. javanica (see Section 4.3). The close relationship between these two, which have previously been considered conspecific (see Section 4.3), and M. cheni, M. passerina and M. williamsi has been suggested based on morphological similarity (de Juana et al., 2004; Wolters, 1979). Clades B1c and B1d contain exclusively African species, and the sister species M. africana and M. hypermetra, as well as M. apiata and M. fasciolata, have been considered to be conspecific or form superspecies (see Section 4.3), so their close associations were expected. In contrast, the predicted close relationship between M. rafuicaminomea/M. anglonias and the M. apiata complex (Dean et al., 1992; de Juana et al., 2004; Pätzold, 2003) is unsupported, and the close association (subgenus Corypha) between these and M. africana and M. hypermetra (and M. somalica and M. sharpii, which were not included here) is only partly supported (M. africana, M. hypermetra, M. apiata and M. fasciolata; clade B1d).

Clades B2a and B2b are both strongly supported (though only cytB and 16S are available for all but one of these species), although all of the relationships within clade B2a except the sister relationship between C. barlowi and C. erythrochlamys are effectively unresolved. The taxonomic history of the taxa in clade B2a is checkered. Two or three of the species C. barlowi and C. erythrochlamys have been treated as conspecific (see Section 4.3), and they have variously been placed in Pätzold, 2003; Peters, 1960), C. burra has been placed in 1992; Dickinson, 2003; Meinertzhagen, 1951; Peters, 2003) or genetic data from Barnes (2007).

Our findings highlight the large number of relationships suggested by molecular data that conflict with previous morphology-based classifications (e.g. Dickinson, 2003; Meinertzhagen, 1951; Pätzold, 2003; Peters, 1960; Sibley and Monroe, 1990; Wolters, 1979; cf. Fig. 3). The treatments by de Juana et al. (2004), Hockey et al. (2005) and Gill and Donsker (2012) are more closely aligned with our findings because they were partly based on mitochondrial data from Barnes (2007) that is only now being published here.

Harrison (1966) suggested, based on a detailed study of morphological characters, that Galerida, Lullula and Pseudalaemon be synonymized with Alauda. At the time, three of the species presently placed in Galerida, i.e. G. deva (not included in the present study), G. magnirostris and G. modesta, were placed in monotypic genera (Spizalauda, Calendulauda and Heliocorys, respectively), and A. razae was placed in a monotypic Spizocorys. The present study supports Harrison’s (1966) proposal only if Spizocorys also is included in Alauda, i.e. the entire clade A2 is referred to as Alauda. However, we prefer to retain Galerida, Alauda, Lullula and Spizocorys. There is no support for upholding the monotypic genus Pseudalaemon, so we synonymize this with Spizocorys. Melanocorypha leucoptera has been considered to form a superspecies with M. mongolica based on plumage similarity and parapatric distributions (Cramp, 1988; Glutz von Blotzheim and Bauer, 1985). However, as the molecular data suggest that M. leucoptera is not closely related to the other Melanocorypha species (including the type species of the genus, M. yeltoniensis), it should be removed from this genus. Its affinity with Alauda is strongly supported in the concatenated analysis, although, as has been pointed out above, this might rest entirely on ODC. As a close relationship with Alauda is indicated also by morphological, vocal, behavioural and ecological data (de Juana et al., 2004; P.A. and Krista Mild, unpublished), we propose that it be treated as Alauda leucoptera.

The non-monophyly of Calandrella is strongly supported by our data. The type species of this genus, C. brachydactyla, is in clade A1d. Accordingly, the species in this clade should retain the generic name Calandrella. For clade A1a, the generic name Alauda belongs to the type species: Calandrella raytal, and we propose that this name be used for the species in this clade, i.e. A. rufescens, A. cheleensis, A. raytal and A. athensis (as was already done by Wolters, 1979; except for the last one, which was placed in the genus Calandrella).
Fig. 3. Morphological variation in some larks. Same tree as in Fig. 1. Different colours of names indicate genera as defined by Peters (1960) based on morphology; monotypic genera are shown in black. Revised names compared to Gill and Donsker (2012) are indicated by *. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
**Mirafr a hova** is firmly anchored in clade C1a, together with *Eremopterix*. Although it is uncertain whether it is sister to all *Eremopterix*, to all *Eremopterix* except *E. australis*, or to *E. australis*, we propose that it be recognised as *Eremopterix hova*.

### 4.3. Taxonomic implications at the species level

Although the main focus of this paper is not on species level taxonomy, some of the results provide important contributions to ongoing debates about species limits, and some reveal previously unknown deep divergences. We do not advocate the use of cut-off values in genetic divergence as taxonomic yardsticks, but instead support an integrative approach based on independent data, whatever species concept is adopted. As dating based on the molecular clock is uncertain (e.g. García-Moreno, 2004; Lovette, 2004; Penny, 2005; but see Weir and Schluter, 2008, whose average rate we have adopted), we emphasise the relative ages of different clades more than the actual ages inferred.

*Guillaumet et al.* (2005, 2006, 2008) discovered two primary clades within *Galerida cristata*, which had reached reciprocal monophyly in mtDNA and showed evidence of strong reproductive isolation in their narrow contact zone in Morocco. These were later recognised as separate species, *Galerida cristata sensu stricto* and *G. macrorhyncha* (Gill and Donsker, 2012). The split between these clades is here estimated to be approximately two thirds of that between the youngest widely sympatric reproductively isolated sister species. As all available *G. macrorhyncha* sequences are from Morocco, at the western edge of the ranged portion of the taxon *randoni* (Cramp, 1988; de Juana et al., 2004), and as there are no samples from or close to the Algerian type localities of *randoni* and *macrorhyncha*, more research is needed on the circumscription and nomenclature of these taxa.

*Guillaumet et al.* (2008) showed using cytb sequences that the subspecies *Galerida theklae praeterrmissa* (Ethiopia) and *G. t. ellioti* (Somalia) are deeply diverged from the northwest African subspecies, and also fairly distinct from each other. Using mainly the same data, the present study infers the split between the populations from northwest Africa and the Horn of Africa to be approximately the same as that between the youngest widely sympatric reproductively isolated species pair. The separation between the two Horn of Africa taxa is inferred to be similar to that between the reproductively isolated, marginally sympatric *G. cristata* and *G. macrorhyncha*. A taxonomic revision is evidently called for, including sequence data for the taxa in the Horn of Africa for which no molecular data are available (*G. t. hararensis*, *G. t. mallasbiensis*, *G. t. huriensis*), and additional data on the Horn of Africa *G. t. hui*, for which a short cytb fragment indicated substantial divergence from *praeterrmis* (Guillaumet et al., 2008).

The taxonomy of the *Calandrella brachyactyla* has been treated as a subspecies of *C. cinerea* (e.g. Meinertzhagen, 1951; Pätzold, 2003; Peters, 1960; Stepanyan, 1990; Vaurie, 1959), but is nowadays usually considered a separate species (e.g. Cramp, 1988; Dean et al., 1992; de Juana et al., 2004; Dickinson, 2003; Gill and Donsker, 2012; Glutz von Blotzheim and Bauer, 1985; Hall and Moreau, 1970; Sibley and Monroe, 1990; Wolters, 1979). Meinertzhagen (1951) included also *C. acutirostris* in *C. cinerea sensu lato*. The results from the present study confirm deep splits between *C. cinerea*, *C. brachyactyla* and *C. acutirostris*, adding further support to the treatment of these as different species. However, completely unexpectedly, they also suggest a deep separation between *C. brachyactyla rubiginosa/C. b. longipennis* from Morocco and Kazakhstan, respectively, and *C. b. dukhunensis* from Mongolia, and strongly support a sister relationship between the latter and *C. acutirostris*. As these results are only based on mitochondrial DNA, a more comprehensive study is needed before any taxonomic revision can be undertaken.

The genus *Eremophilia* comprises only two species, *Eremophilia bilopha* is restricted to North Africa and the Middle East, whereas *E. alpestris* is the most widely distributed of all lark species, breeding on five continents, and is the only lark native to the New World (de Juana et al., 2004). Morphological variation is pronounced in *E. alpestris*, with 40–42 subspecies recognised (de Juana et al., 2004; Peters, 1960). The present study includes just a small portion of this variation, but nevertheless indicates that *E. alpestris* is probably better treated as multiple species. That our sample of the Central Asian *E. a. brandti* is inferred to be more closely related to the two North American samples than to the other Eurasian taxa is totally unexpected, and requires confirmation. If corroborated by independent data, this implies a complex biogeographical history for this species group.

The widespread *M. cantillans*, which ranges from west Africa to India, and the similarly widely distributed *M. javanica*, from Myanmar to Australia (de Juana et al., 2004) have previously been considered conspecific (Dickinson and Dekker, 2001; Pätzold, 2003; Peters, 1960; Vaurie, 1951; reviewed in first reference). The close relationship between these two is confirmed by the present study. Both species are monophyletic in the cytb tree, although their separation is comparatively recent (1.2 MYA; 0.7–1.7 MYA, 95% HPD), only slightly more than one third of the age of the youngest widely sympatric species pair. These taxa have apparently spread over a vast area in a very short time, and are in the early stages of the speciation process. Although the extended cytb tree suggests that they are independently evolving lineages, additional sampling might reveal incomplete sorting of haplotypes, and the ODC sequences do not sort according to species. Independent data are needed to corroborate our results.

*Mirafr a affinis*, *M. erythrocephala* and *M. microptera* were previously treated as subspecies of *Mirafr a assamica* (reviews in Alström, 1998; Dickinson and Dekker, 2001). Alström (1998) proposed that these four (using the name *M. mariona* for *M. erythrocephala*) were better treated as separate species, based on pronounced differences in especially vocalisations and display-flights. This is corroborated by the evidence presented here (and has been accepted by most recent authors, e.g. de Juana et al., 2004; Dickinson, 2003; Gill and Donsker, 2012). Although the relationships among these species are largely unsupported, our data suggest that *M. erythroptera* is nested within the *M. assamica* complex, and that *M. microptera* is sister to the others. The splits among these species are inferred to be at least twice as old as the oldest widely sympatric sister pair in the entire study.

*Mirafr a apiana* and *M. fasciolata* were traditionally treated as conspecific (e.g. Dean et al., 1992; Pätzold, 2003; Peters, 1960; Wolters, 1979), but have recently been suggested to be separate species (de Juana et al., 2004; Hockey et al., 2005) based on limited
unpublished genetic data. The present study confirms that these two taxa have been separated for a long time.

Calendulauda albescens, C. barlowi and C. erythrochlamys have been treated as conspecific (under the first name; Peters, 1960; Wolters, 1979), or C. erythrochlamys has been split off as a separate species (Dean et al., 1992; Sibley and Monroe, 1990). Ryan et al. (1998) suggested, based on a study of cytb, morphology and song, that three species should be recognised, and this has been followed by most subsequent authors (Dickinson, 2003; de Juana et al., 2004; Gill and Donkers, 2012; Hockey et al., 2005). The relationships among these are uncertain, as cytb and 16S support different topologies in relation to C. burra. The extended cytb dataset suggests deep splits among C. albescens, C. burra and C. barlowi/C. erythrochlamys, considerably older than the split between the widely sympatric Melanocorypha maxima and M. mongolica, adding further support to the treatment of these as separate species. However, the divergence between C. barlowi and C. erythrochlamys is the second most recent of all pairs treated as different species. Accordingly, in the absence of other data, whether C. barlowi should be given specific status or treated as a subspecies of C. erythrochlamys (by priority) is an open question. The same applies to C. alapex, which is often considered a subspecies of C. africanoides (e.g. Dean et al., 1992; Pätzold, 2003; Peters, 1960), although the divergence between these two is slightly deeper than between C. barlowi and C. erythrochlamys.

Ammomanes deserti is widely distributed across North Africa to western India, with 23–24 subspecies recognised (de Juana et al., 2004; Peters, 1960). Although the present study only covers a tiny fraction of the geographical variation, it nevertheless infers four deeply-diverging cytb lineages, suggesting that A. deserti is in need of further study and taxonomic revision. Additionally, A. cinctura, which occurs from the Cape Verde islands through North Africa to southwest Pakistan, with three subspecies recognised (de Juana et al., 2004; Peters, 1960) shows an unexpected deep cytb divergence between samples of the same subspecies (arencolor) from Morocco and Saudi Arabia. More extensive sampling of this species also is warranted.

Five Certhilauda species (all except C. chuana) previously have been treated as conspecific under the name C. curvieri (Meinertzhagen, 1951; Pätzold, 2003; Peters, 1960; Wolters, 1979), although they have recently been split based on differences in mitochondrial DNA (Ryan and Bloomer, 1999; followed by Dickinson, 2003; de Juana et al., 2004; Gill and Donksker, 2012; Hockey et al., 2005). The divergence between C. subcoronata and C. benguelensis is substantial (despite limited morphological differentiation), as is the divergence between these two and the three other species in this complex. In contrast, the separation between C. brevirostris, C. semitorquata and C. curvieri is much more recent. Divergence between the two former taxa is the shallowest of all taxa currently treated as different species, yet they have divergent ranges, separated by a population of C. subcoronata. These three taxa are in the early stages of the speciation process, and their taxonomic ranking is therefore open to different interpretations.

The two Chersomanes species (C2b) were previously often considered conspecific (Dean et al., 1992; Pätzold, 2003), but were separated by de Juana et al. (2004) based on unpublished genetic differences, widely disjunct distributions and differences in sexual plumage dimorphism (slight in beesleyi, absent in albifasciata). This separation has since been questioned (Donald and Collar, 2011), but the present study confirms their long separation, adding further support to their treatment as separate species (although better coverage of northern populations of albifasciata is desirable).

4.4. Strongly heterogeneous morphological evolution

Larks provide extraordinary examples of the effects of natural selection on phenotypes, and few groups of birds show the same level of disagreement between taxonomy, based on morphology, and phylogenetic relationships as inferred by DNA. Although the present study does not examine morphological divergence quantitatively, it nevertheless indicates multiple examples of highly conserved phenotypes as well as dramatic morphological divergence in certain lineages and instances of parallel evolution (Fig. 1). Traits related to feeding, such as size and shape of bill, appear to be particularly labile, with striking differences between some sister species as well as, conversely, close similarities among distantly related species. For larks, which inhabit mostly open habitats, cryptic plumages are evidently important. Consequently, the strength of streaking and colour shades above appear to be particularly adaptable, reflecting the amount of vegetation cover (aridity) and substrate colour more than phylogeny.

The similarities in size, structure and plumage between the two distantly related clades of traditional Calandrella (here recognised as Calandrella and Alauda; cf. de Juana et al., 2004; Fig. 3) are likely the result of either retained plesiomorphies or parallel evolution. The similarity between the north African/west Asian Eremalauda durni and Afrotropical Spizocorys starki, between the Western Palearctic Chersophilus and Afrotropical Certhilauda, and between the north African/west Asian Ammannopodini and Afrotropical Ammanomanopsis (cf. de Juana et al., 2004; Fig. 3) provide examples of close morphological similarity evolving independently in similar environments. In contrast, the dissimilarity between Ammanomanopsis and its closest relatives, Chersomanes and Certhilauda, suggests strong divergence in the former. The sister relationship between the genera Calandrella (as redefined here) and Eremophila suggests remarkable plumage divergence in the latter lineage (which is one of the most aberrant of all larks; cf. de Juana et al., 2004 and Fig. 3). Similarly, the close relationship between Alauda (as redefined here; clade A1a) and the two monotypic genera Eremalauda and Chersophilus reveal extraordinary changes in both structure (especially bill) and plumage among sister taxa (cf. de Juana et al., 2004; Fig. 3). Meinertzhagen's (1951) inappropriate placement of Chersophilus, Pseudalaemon, Calendulauda, Alaeo, Chersomanes and Certhilauda in one genus based on bill structure and behaviour (notably strong digging with the bill when feeding, and fast running) is a striking example of a misclassification caused by the strong lability and adaptability of bill morphology in larks.

Within the true Melanocorypha clade (A1c), there is much variation, especially with respect to plumage (cf. de Juana et al., 2004; Fig. 3). M. yeltoniensis is one of the few larks with pronounced sexual dimorphism in plumage: females have cryptic, plesiomorphic, plumages reminiscent of M. bimaculata and M. calandra, whereas males are practically all black in the breeding season (somewhat more cryptic in the non-breeding season); also the size differences between females and males are pronounced. The plumage similarity between M. mongolica and Alauda leucophris (previously M. mongolica), which has been assumed to be due to close relationship (e.g. Pätzold, 2003; Wolters, 1979) is apparently due to parallel evolution.

Apart from Melanocorypha yeltoniensis, the sparrow-larks Eremopterix spp. are the only larks with strong sexual plumage dimorphism, and the male plumages are contrastingly patterned in black and white on the head and underparts, except in E. australis, which lacks white (cf. de Juana et al., 2004; Fig. 3). However, the strongly supported inclusion of the Madagascar endemic Mirafra hova in this clade, and hence its suggested transfer to Eremopterix, is most remarkable in view of its strikingly different plumage from all plumages of other Eremopterix species and close similarity to some Mirafra species (cf. de Juana et al., 2004; Fig. 3). The uncertainty regarding its position in the tree in relation to E. australis (and hence also the other Eremopterix species) precludes reconstruction of the evolution of sexual dimorphism and typical male Eremopterix plumage.
Apart from the species with strong sexual dimorphism in plumage, Melanocorypha yeltoniensis and the sparrow-larks Eremonterix spp. (except E. hova), slight plumage differences between the sexes is present in Eremonterix spp., Alauda leucoptera, Ramphocoris clotebay and Pinarcorys erythropygia (de Juana et al., 2004), showing that sexual plumage dimorphism has evolved multiple times.

The molecular data suggest that the similarities between Gale-risidae theklae and G. malabarica, which have often been treated as conspecific (e.g. Dean et al., 1992; Hall and Moreau, 1970; Howard and Moore, 1994), are due to parallel evolution, although retention of plesiomorphies cannot be eliminated based on the available data. In contrast, the divergent morphology of the Cape Verde endemic Alauda razeae (not shown) compared to the other species of Alauda (cf. de Juana et al., 2004) has misled earlier workers regarding its generic affinities (Boyd Alexander, 1898; Meierzhagen, 1951; Pützold, 2003; Peters, 1960; Vaurie, 1959; Voous, 1977; Wolters, 1979). This disparity agrees with the rapid morphological evolution typical of many small island populations (Grant, 1998).

Within the Spizocorys clade there is considerable variation (cf. de Juana et al., 2004; Fig. 3), especially with respect to pigmentation, head pattern (notably S. personata) and bill size/shape (especially S. fremantili), which has confused earlier taxonomists. The morphological similarity between Spizocorys and Calandrella (which led Meierzhagen, 1951, to unite these genera) is apparently the result of parallel evolution. Conversely, based on morphology (cf. de Juana et al., 2004; Fig. 3), the close relationship between Spizocorys and the monotypic Lullula is totally unexpected. Similarly, the close relationship between Ramphocoris, Pinarcorys and Ammomanes is highly surprising when viewed from a purely morphological perspective; in particular the bill morphology of Ramphocoris is unique among the larks (cf. de Juana et al., 2004; Fig. 3).

In the Mirafr/a/Heteromirafr/a clade (B1), plumage variation mainly concerns colour tones and strength of streaking, whereas the variation in bill morphology is more pronounced (cf. de Juana et al., 2004; Fig. 3). Morphological divergence has apparently been extremely slow over substantial time periods in some clades, e.g. in the five species in the M. assamica–M. erythropotera complex (clade B1a), which until recently was usually treated as two species, but which was here inferred to have been separated for millions of years. Conversely, in the closely related Calendulauda clade (B2), the variation in plumage and structure is more pronounced (cf. de Juana et al., 2004; Fig. 3) that the species placed in this genus have previously been placed in five different genera. Even within clade B2a, the variation in plumage and bill size is marked.

5. Conclusions

Our analyses support the contention that incomplete data sets, especially those where one or a few loci have been consistently sampled from all taxa, can provide robust, well-resolved hypothe-ses of relationships (Wiens et al., 2005; Wiens and Morrill, 2011; but see Lemmon et al., 2009). Overall, our concatenated tree shows little conflict with individual gene trees, but a few specific relationships do show evidence of conflict, possibly due to differential lineage sorting. This highlights the continued importance of performing single gene as well as combined data analyses, since the latter may obscure significant incongruence behind strong nodal support values. The multilocus tree inferred here revealed many unpredicted relationships, including some non-monophyletic genera. The dated cyt b tree indicated some unexpectedly deep divergences between taxa currently regarded as subspecies and one non-monophyletic species, as well as some comparatively shallow splits between currently recognised species. The phylogeny indicates multiple examples of parallel morphological evolution, probably resulting from variation in selective forces (both natural and sexual) associated with the broad array of open habitats where larks occur. In contrast to the overall rather conserved plumage evolution in larks, some close relatives show dramatic differences in plumage and bill structure, with the latter appearing to be particularly labile. Future work should focus on quantifying rates of evolution in these traits in the context of our robust phylogenetic framework. Few groups of birds show the same level of disagreement between morphologically-based taxonomy and phylogenetic relationships as inferred using DNA data.

Acknowledgments

We are most grateful to the following colleagues and institutions for providing samples and/or for assistance in the field: Raúl Aymí, Neil and Liz Baker, John Bates, Oleg Belyalov, Geoff Carey, Callan Cohen, Nigel Collar, Miles Coverdale, Edward Gavrilo, Andrew Griewe, Cornelis Hazevoet, Daniel Hooper, Nick Horrocks, Björn Johansson, Joris Komen, Andrew Lassey, Paul Leader, Dave Moyer, H. Nikolaus, Trevor Price, Hadoram Shirihai, Claire Spottiswoode, Martin Stervander, Lars Svensson, Irene Tieleman, Per Undeland, Jo Williams, Bill Zetterström; Leon Bunnun and George Amutete and the Kenyan National Museum; Leo Joseph and the Australian National Wildlife Collection; Michel Louette and the Royal Museum for Central Africa, Tervuren; Göran Frisk, Ulf Johansson and Peter Nilsson and the Swedish Museum of Natural History; Silke Fregin and Martin Haase and the Vogelwarte Hiddensee, Ernst Moritz Arndt University of Greifswald; Jan Lifeld and the National Centre for Biosystematics, Natural History Museum, Oslo; John Bates and the Field Museum of Natural History, Chicago: Jean-Marc Pons and the Muséum National d’Histoire Naturelle, Paris; and Sharon Birks and the University of Washington Burke Museum. Our sincerest thanks also to Wayne Delport, Lisel Solms and Isa-Rita Russo for assistance in the lab, and to Shimiao Shao for submitting the sequences to GenBank. We also thank Josep del Hoyo, Andrew Elliott and Lynx Edicions for permitting the use of paintings of larks from the Handbook of the Birds of the World, and to Anders Rådén for assistance in their scanning. We are indebted to Jornvall Foundation, Riksmusei Vanners Linnaeus award and the Chinese Academy of Sciences Visiting Professorship for Senior International Scientists (No. 2011T2S04) (all to P.A.); to the Pakistan Science Foundation (Research Project No. PSF/Res./P-BZU/Bio(340); to A.A.K and M.A.Q.); and to University of Cape Town and South African National Research Foundation (to P.B. and P.G.R.). Collections in Morocco were facilitated and funded in part by the International Foundation for the Development and Conservation of Wildlife (IFCDW). We are most grateful to Margaret Koopman at the Niven Library, Percy FitzPatrick Institute of African Ornithology, for locating important references; to Krister Mild and Jan Sundler for various assistance; to Normand David for his expert advice on the ending of the name Eremonterix hova; and to two anonymous reviewers for comments on the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.06.005.

References
