Zoological Journal of the Linnean Society, 2023, **XX**, 1–29
https://doi.org/10.1093/zoolinnean/zlad107
Advance access publication 28 September 2023
Original Article

Original Article

Integrative taxonomy reveals unrecognised species diversity in African *Corypha* larks (Aves: Alaudidae)

Per Alström[1,2,*], Zeinolabedin Mohammadi[3], Paul F. Donald[4,5], Marianne Nymark[1], Erik D. Enbody[6,7], Martin Irestedt[8], Emmanuel Barde Elisha[9,10], Henry K. Ndithia[11], B. Irene Tieleman[12], Derek Engelbrecht[13], Urban Olsson[14,15], Loïs Rancilhac[1], Martin Stervander[16,17,*]

[1]Animal Ecology, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden
[2]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, Chinaa
[3]Department of Biology, Faculty of Sciences, Golestan University, Gorgan, Iran
[4]BirdLife International, The David Attenborough Building, Pembroke Street, Cambridge CB2 3QZ, UK
[5]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
[6]Department of Medical Biochemistry and Microbiology, Uppsala University, 751 23 Uppsala, Sweden
[7]Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA 95060, USA
[8]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, PO Box 50007, 104 05 Stockholm, Sweden
[9]Nigerian Montane Forest Project (NMPF), Ngel Nyaki Forest Reserve, Mambilla Plateau, Taraba State, Nigeria
[10]A.P. Leventis Ornithological Research Institute (APLORI), Biological Conservatory, Jos-East, Plateau State, Nigeria
[11]Department of Zoology, Ornithology Section, National Museums of Kenya, PO Box 40658-00100 GPO, Nairobi, Kenya
[12]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands
[13]Department of Biodiversity, University of Limpopo, Private Bag X1106, Sovenga, 0727, South Africa
[14]Department of Biological and Environmental Sciences, University of Gothenburg, PO Box 463, 405 30 Göteborg, Sweden
[15]Gothenburg Global Biodiversity Centre, PO Box 461, 405 30 Göteborg, Sweden
[16]Bird Group, Natural History Museum, Akeman Street, Tring, Hertfordshire HP23 6AP, UK
[17]Present address: Department of Natural Sciences, National Museums Scotland, Chambers Street, Edinburgh EH1 1JF, UK

[*]Corresponding authors. Animal Ecology, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden. E-mail: per.alstrom@ebc.uu.se; Bird Group, Natural History Museum, Akeman Street, Tring, Hertfordshire HP23 6AP, UK. E-mail: martin@stervander.com

ABSTRACT

The species complex comprising the rufous-naped lark *Corypha africana*, Sharpe's lark *Corypha sharpii*, the red-winged lark *Corypha hypermetra*, the Somali long-billed lark *Corypha somalica* and Ash's lark *Corypha ashi* encompasses 31 recognised taxa across sub-Saharan Africa, many of which are extremely poorly known and some not observed for decades. Only 17 taxa have been studied molecularly and none comprehensively for morphology, vocalisations or other behaviours. Here, we undertake comprehensive integrative taxonomic analyses based on plumage and morphometrics (for 97% of the taxa), mitochondrial and nuclear loci (77%), ≤ 1.3 million genome-wide single nucleotide polymorphisms (68%), song (many described for the first time; 52%) and additional behavioural data (45%). All polytypic species as presently circumscribed are paraphyletic, with eight primary clades separated by ≤ 6.3–6.8 Myr, broadly supported by plumage, morphometrics, song and other behaviours. The most recent divergences concern sympatric taxon pairs usually treated as separate species, whereas the divergence of all clades including *C. africana* subspecies is as old as sister species pairs in other lark genera. We propose the recognition of nine instead of five species, while *C. ashi* is synonymised with *C. somalica rochei* as *C. s. ashi*. The geographical distributions are incompletely known, and although the nine species are generally para-/allopatric, some might be sympatric.

Keywords: Africa; behaviour; bird; morphometrics; new classification

Received 26 February 2023; revised 24 July 2023; accepted 26 July 2023

INTRODUCTION

The avian family Alaudidae (larks) comprises ~100 species, which are widely distributed across Eurasia and Africa, with one species ranging to Australia and one to North America and Colombia (de Juana *et al.* 2004, Fjeldså *et al.* 2020a, Alström *et al.* 2023, Gill *et al.* 2023). Larks generally inhabit open areas, such as various types of grasslands, open scrubland and deserts. This habitat choice is exceptional in comparison to the other ~1,100 species within the large superfamily Sylvioidea (*sensu* Fregin *et al.* 2012, Alström *et al.* 2006, Fregin *et al.* 2012, Fjeldså *et al.* 2020a; reviewed by Alström *et al.* 2013b, Fjeldså *et al.* 2020a).

Phylogenetic relationships within the Alaudidae were reconstructed a decade ago using mitochondrial markers and a few nuclear markers for ~80% of the species, revealing an extremely high discordance between traditional morphology-based taxonomy and the phylogeny (Alström *et al.* 2013a). These findings were confirmed by a recent study using a genus-level sampling of genome-wide single nucleotide polymorphisms (SNPs) coupled with ≤ 17 loci in all but one species (Alström *et al.* 2023). As a consequence, a new subfamily classification was proposed (with subfamilies Certhilaudinae, Mirafrinae and Alaudinae), as well as a division of the genus *Mirafra* into four genera (*Mirafra s.s.*, *Corypha*, *Amirafra* and *Plocealauda*) to balance the generic classification of the Alaudidae (Alström *et al.* 2023).

Most larks display different shades of brown on the upperparts and paler underparts, with variously prominent streaking above and below, rendering many species extremely similar in appearance (de Juana *et al.* 2004). In addition, many larks show an inordinately strong association between plumage coloration and the substrates on which they occur (Donald *et al.* 2017). The songs vary from extremely complex and varied, often including mimicry, to very simple and repetitive (de Juana *et al.* 2004, Shirihai and Svensson 2018). Some species have very similar songs, even when sympatric, whereas some closely related species have strikingly different songs. Most species undertake song flights, which can be of short duration and at low height or lasting for long periods and performed at great height (de Juana *et al.* 2004).

Recent years have seen a surge in the number of recognised or proposed lark species, owing to studies using, in particular, molecular markers and bioacoustics (Table 1). Although the taxonomy of Eurasian larks is relatively well studied, relatively fewer studies have dealt with African species, despite the fact that 60% of the lark species occur exclusively in Africa (Fjeldså *et al.* 2020a). The newly resurrected genus *Corypha*, which was until recently treated as part of a very large genus, *Mirafra* (Alström *et al.* 2023), contains a complex of superficially similar but poorly known species comprising the following: the rufous-naped lark *Corypha africana* (A. Smith, 1836), with 22–23 subspecies; the red-winged lark *Corypha hypermetra* (Reichenow, 1879), with four subspecies; the Somali long-billed lark *Corypha somalica* (Witherby, 1903), with two subspecies; and the monotypic Ash's lark *Corypha ashi* (Colston, 1982) (Clements *et al.* 2022, Gill *et al.* 2023; Table 2). A fifth species, the monotypic Sharpe's lark, *Corypha sharpii* (Elliot, 1897), has recently been recognised (del Hoyo and Collar 2016, Clements *et al.* 2022, Gill *et al.* 2023), although it is often treated as a subspecies of *C. africana* (Dickison & Remsen, 2014; Gill *et al.* 2023). This species complex (henceforth referred to as the *Corypha* complex)

Table 1. Integrative taxonomic studies of larks and the types of data used.

Taxa	Geographical region	M	V	B	E	MtDNA	NcDNA	References
Plocealauda assamica (Horsfield, 1840) complex[a]	South Asia	x	x	x	x	x		Alström (1998), Alström *et al.* (2013a)
Calendulauda albescens (Lafresnaye, 1839) complex	Sub-Saharan Africa	x	x			x		Ryan *et al.* (1998)
Certhilauda curvirostris (Hermann, 1783) complex	Sub-Saharan Africa		(x)			x		Ryan and Bloomer (1999)
Galerida cristata (Linnaeus, 1758)–*Galerida macrorhyncha* Tristram, 1859 complex	Palaearctic	x			x	x		Guillaumet *et al.* (2005), Guillaumet *et al.* (2006), Guillaumet *et al.* (2008)
Calandrella brachydactyla (Leisler, 1814)–*Calandrella dukhunensis* (Sykes, 1832) complex	Palaearctic	x	x	x		x	x	Alström *et al.* (2013a), Stervander *et al.* (2016), Alström and Sundev (2021)
Eremophila Boie, 1828 complex	Palaearctic	(x)				x	(x)	Drovetski *et al.* (2014), Ghorbani *et al.* (2020a)
Alaudala rufescens (Vieillot, 1819)–*Alaudala cheleensis* Swinhoe, 1871–*Alaudala raytal* (Blyth, 1845) complex	Palaearctic	x	x	x	x	x	x	Ghorbani *et al.* (2020b), Alström *et al.* (2021)
Calandrella blanfordi (Shelley, 1902)–*Calandrella erlangeri* (Neumann, 1906) complex	Sub-Saharan Africa	x				x	(x)	Stervander *et al.* (2016)
Heteromirafra Grant, 1913 complex	Sub-Saharan Africa	x	x			x		Spottiswoode *et al.* (2013)

Abbreviations: B, behaviour; E, ecology (including habitat); M, morphology; mtDNA, mitochondrial DNA; ncDNA, nuclear multilocus or genome-wide DNA; V, vocalisations.
[a]*Mirafra assamica* in original publication.

Table 2. Distributions, type localities, samples per data category and proposed new classification of the taxa in the *Corypha africana–C. sharpii–C. hypermetra–C. somalica–C. ashi* complex.

Taxon (Clements *et al.* 2022)	Distribution	Type locality	SNP	Mloc	Morph	Song	Beh	Proposed new classification	English name (proposed or in use)
Rufous-naped lark *Corypha africana* (Smith, A, 1836)									
C. a. africana (Smith, A, 1836)	South-east South Africa	South Africa, east province of the Colony as far as Latakoo > Algoa Bay, South Africa (33°58′S, 25°38′E)	x	x	x	x	x	***Corypha africana*** *africana*	Rufous-naped lark
C. a. transvaalensis (Hartert, E, 1900)	Tanzania–north South Africa	Rustenburg, Transvaal, South Africa (25°40′S, 27°15′E)		x	x	x	x	*C. africana africana*	
C. a. grisescens (Sharpe, 1902)	West Zambia, north Botswana, north-west Zimbabwe	Zimbabwe Tibakai's Vley, Matabele Land = Tibakai's kraal, Gabzuma Pan, Wankie (18°54′S, 25°56′E)	x	x	x	x	x	*C. africana grisescens*	
C. a. pallida (Sharpe, 1902)	South-west Angola, north-west Namibia	Namibia, Damara Land > Elephant Vley, (Ovamboland) (17°59′S, 17°32′E)			x	x	x	*C. africana pallida*	
C. a. ghansiensis (Roberts, 1932)[a]	East Namibia, west Botswana	Botswana Gemsbok Pan (21°43′S, 21°38′E)			x			*C. africana ghansiensis*	
C. a. chapini (Grant, CHB & Mackworth-Praed, 1939)	South-east DR Congo, north-west Zambia	Zaire Kanzenze (= Nasondoye) (10°22′S, 25°04′E)		x	x			*C. africana chapini*	
C. a. occidentalis (Hartlaub, 1857)	West Angola	Gaboon [sic] > Angola, Catumbella (12°25′S, 13°32′E)		x	x	–	–	*C. africana occidentalis*	
C. a. gomesi (White, CMN, 1944)	East Angola, west Zambia	Angola east of Lunyuwe R., Macondo district (~12°30′S, 23°30′E)			x			*C. africana gomesi*[d]	
C. a. isolata (Clancey 1956)[b]	South-east Malawi	Malawi, 15 miles north-west of Namwera, Fort Johnston district, southern end of Lake Nyasa, Nyasaland (14°22′S, 35°30′E)						*C. africana isolata*[a]	
C. a. tropicalis (Hartert, E, 1900)	East Uganda, west Kenya–north-west Tanzania	Tanzania, Tropical East Africa to Lake districts and Uganda > Tanzania Bukoba (1°19′S, 31°49′E)	x	x	x	x	x	*C. africana tropicalis*	
C. a. ruwenzoria (Kinnear, 1921)	East DR Congo–south-west Uganda	Uganda Mokia river (0°06′N, 30°03′E)			x	x	x	*C. africana tropicalis*	
C. a. kurrae (Lynes, 1923)	West Sudan	Sudan, Kurra, Darfur (13°16′N, 24°30′E)	x	x	x			***C. kurrae*** *kurrae*	Highland lark
C. a. batesi (Bannerman, 1923)	Central Nigeria–south-east Niger, west Chad	Nigeria near Jos (9°54′N, 8°53′E)			x			*C. kurrae batesi*	

Table 2. Continued

Taxon (Clements et al. 2022)	Distribution	Type locality	SNP	Mloc	Morph	Song	Beh	Proposed new classification	English name (proposed or in use)
C. a. stresemanni (Bannerman, 1923)	North-central Cameroon	Cameroon, 15 miles north of Ngaundere (7°19′N, 13°35′E)	x	x	x	x	x	*C. kurrae stresemanni*	
C. a. henrici (Bates, GL, 1930)	Guinea–south-west Ivory Coast	Guinea, Mt Nimba > Bossu, Mt Nimba (7°38′N, 8°30′W)	x	x	x			*C. kurrae henrici*	
C. a. bamendae (Serle, 1959)	West Cameroon	Cameroon, Sabga Pass, Bamenda (6°01′N, 10°19′E)	x	x	x	x		*C. kurrae bamendae*	
C. a. athi (Hartert, E, 1900)	Central Kenya–north-east Tanzania	Kenya, Athi Plain 1°27′S, 36°59′E	x	x	x	x	x	***C. athi*** *athi*	Sentinel lark
C. a. harterti (Neumann, 1908)	South-central Kenya	Kenya, Kiboko river, south Ukamba (2°11′S, 37°43′E)	x	x	x	x		*C. athi harterti*	
C. a. kabalii (White, CMN, 1943)	North-east Angola, north-west Zambia	Zambia, Minyanya Plain, western Balovale (13°09′S, 22°23′E)	x	x	x	x	x	***C. kabalii*** *kabalii*	Plains lark
C. a. irwini (da Rosa Pinto, 1968)[c]	South-east Angola	Angola, Rio Longa, Cuando-Cubango (14°42′S, 18°30′E)	x	x	x			*C. kabalii irwini*	
C. a. malbranti (Chapin, 1946)	Gabon–south DR Congo	Congo, 30 km south of Djambala (2°30′S, 14°45′E)	x	x	x	x		*C. kabalii malbranti*	
C. a. nigrescens (Reichenow, 1900)	North-east Zambia, south Tanzania	Tanzania, Elton Pass, north of Kondeland > Kitulo Plateau (9°05′S, 33°55′E)	x	x	x	x		***C. nigrescens*** *nigrescens*	Plateau lark
C. a. nyikae (Benson, 1939)	East Zambia, north Malawi, south-west Tanzania	Malawi, Nyika Plateau (10°40′S, 33°51′E)	x	x	x	x	x	*C. nigrescens nyikae*	
Sharpe's lark *C. sharpii* (Elliot, DG, 1897)	North-west Somalia	Somalia, Silo Plain > Bannaanka Seelaley (~9°30′N, 44°30′E)	x	x	x			***C. sharpii***	Russet lark
Red-winged Lark *Corypha hypermetra* (Reichenow, 1879)									
C. h. hypermetra (Reichenow, 1879)	South Somalia–north-east Tanzania	Kenya, Kibaradja, Tana river (1°42′S, 40°07′E)	x	x	x	x	x	***C. hypermetra*** *hypermetra*	Red-winged lark
C. h. gallarum (Hartert, E, 1907)	Ethiopia	Ethiopia, Bouta (i.e. Bonta), Hawash valley (9°25′N, 40°15′E)	x	x	x	x	x	*C. hypermetra gallarum*	
C. h. kidepoensis (Macdonald, 1940)	South Sudan, north-east Uganda	Sudan, Ero, Didinga Mts. (4°13′N, 33°48′E) [given as 4°6′N, long. 33°44′E by Peters]	#	§	x	x	x	***C. kidepoensis*** *kidepoensis*	Kidepo lark
C. h. kathangorensis (Cave, 1940)	South-east Sudan	Sudan, Kathangor (5°45′N, 33°59′E) [Given as 5°45′N, 33°59′E by Peters]			x			*C. kidepoensis kathangorensis*	

Table 2. Continued

Taxon (Clements *et al.* 2022)	Distribution	Type locality	SNP	Mloc	Morph	Song	Beh	Proposed new classification	English name (proposed or in use)
Somali lark *Corypha somalica* (Witherby, 1903)									
C. s. somalica (Witherby, 1903)	North Somalia	Somalia Dibbit (~6°50′N, 47°30′E)	x	x	x	x	x	***C. somalica somalica***	Somali lark
C. s. rochei (Colston, 1982)	Central Somalia	Somalia, 24 km north-east of Uarsciek (= Warsheikh) (2°17′N, 45°50′E)	x	x	x			*C. somalica ashi*	
Ash's lark *Corypha ashi* (Colston, 1982)	East Somalia	Somalia, 13 km north of Uarsciek (= Warsheikh) (2°17′N, 45°50′E)	x	x	x			*C. somalica ashi*	

In some cases, the original type locality has been restricted or amended, which then follows '>'; '#' indicates that only low-coverage SNP data were analysed; '§' indicates that only the mitochondrial *Cytb* was analysed. The original taxonomy and distributions follow Clements *et al.* (2022). Data categories are as follows: Beh, behaviour; Mloc, sequence multilocus; Morph, morphometrics; SNP, genomic SNPs.

[a]Explicitly synonymized with *pallida* by Clancey (1956).

[b]Synonymized with *transvaalensis* by Peters (1960) without further comments. Recognized by Dickinson & Christidis (2014) with comment 'For recognition see Dean *et al.* (1992)'; however, we find no comments supporting this taxon in this reference (cited here as Dean and Keith 1992).

[c]Not recognized by Gill *et al.* (2023).

[d]Only photographs of a single specimen (AMNH347435) seen, but appears possibly to be a junior synonym of *chapini*, although White, in the type description, states that it is 'totally different' from *chapini*.

[e]Only tentatively recognized, because no material has been examined.

occurs more or less patchily in open savannah-type habitats across sub-Saharan Africa (Fig. 1). The species are medium-sized to large for larks (lengths, 16–23 cm; weights, 33–68 g), with streaked upperparts and breasts and with variably prominent rufous panels on their remiges (Ryan 2020a, 2020b, 2020c, Ryan *et al.* 2020). Until recently, only three of the taxa in this complex had been analysed phylogenetically for the mitochondrial cytochrome *b* (*Cytb*) gene (Stervander *et al.* 2020) and two of these also by three nuclear introns (Alström *et al.* 2013a), while a recent study included multilocus data for 17 of them (Alström *et al.* 2023). The song and other behaviours of *C. africana* are usually described briefly in the literature, without reference to subspecies (e.g. Dean 2005, Ryan *et al.* 2020), and the songs and other behaviours are poorly described for *C. somalica* (Ryan 2020c) or undescribed for *C. sharpii* and *C. ashi*. The song of *C. hypermetra* has been described in some detail by Dowsett-Lemaire and Dowsett (1978) and Ryan (2020a). All species are poorly known with respect to their plumage variation, vocalisations, behaviour, distribution and other aspects of their natural history, and some taxa have not been observed in the field for decades. Alström *et al.* (2023) suggested, based on analyses of mitochondrial markers and a few nuclear markers, that nine species should be recognised in this complex: *C. somalica* (with *C. ashi* treated as synonymous with *C. somalica rochei*, under the name *C. somalica ashi*), *C. hypermetra*, *C. sharpii*, *C. kidepoensis* (Macdonald, 1940), *C. athi* (Hartert, EJO, 1900), *C. africana* (*s.s.*), *C. nigrescens* (Reichenow, 1900), *C. kabalii* (White, CMN, 1943) and *C. kurrae* (Lynes, 1923).

Here, we revise the taxonomy of the *Corypha* lark complex by integrating multiple genetic and phenotypic datasets across 31 taxa: plumage and morphometrics for 97%, mitochondrial loci and three nuclear loci for 77%, ≤ 1.3 million genome-wide SNPs for 68%, song for 52% and additional behavioural data for 45% of the taxa; all these traits were analysed for representatives for all nine species proposed by Alström *et al.* (2023), except song and behaviour for *C. sharpii*. This work includes the first extensive genomic study of this group and the first descriptions of songs and details of 'wing-clapping' behaviour for multiple taxa, in addition to the most comprehensive study to date of their geographical distributions.

MATERIALS AND METHODS

Terminology

We follow Alström *et al.* (2023) in subdividing the traditional large genus *Mirafra* into four genera, placing the focal species of the present study in the genus *Corypha* [together with two southern African species, the Cape clapper lark *Corypha apiata* (Vieillot, 1816) and eastern clapper lark *Corypha fasciolata* (Sundevall, 1850)]. Our baseline species/subspecies-level taxonomy follows Clements *et al.* (2022), except that we also recognise *Corypha africana irwini*; in total, 31 taxa (Table 2). Binomial names are followed by '*s.l.*' when referring to their traditional definition and '*s.s.*' when referring to less inclusive species delimitations proposed here. When a taxon name is not preceded by the genus name (e.g. *africana*, *sharpii*), we refer to the least-inclusive described taxon (either subspecies or monotypic species).

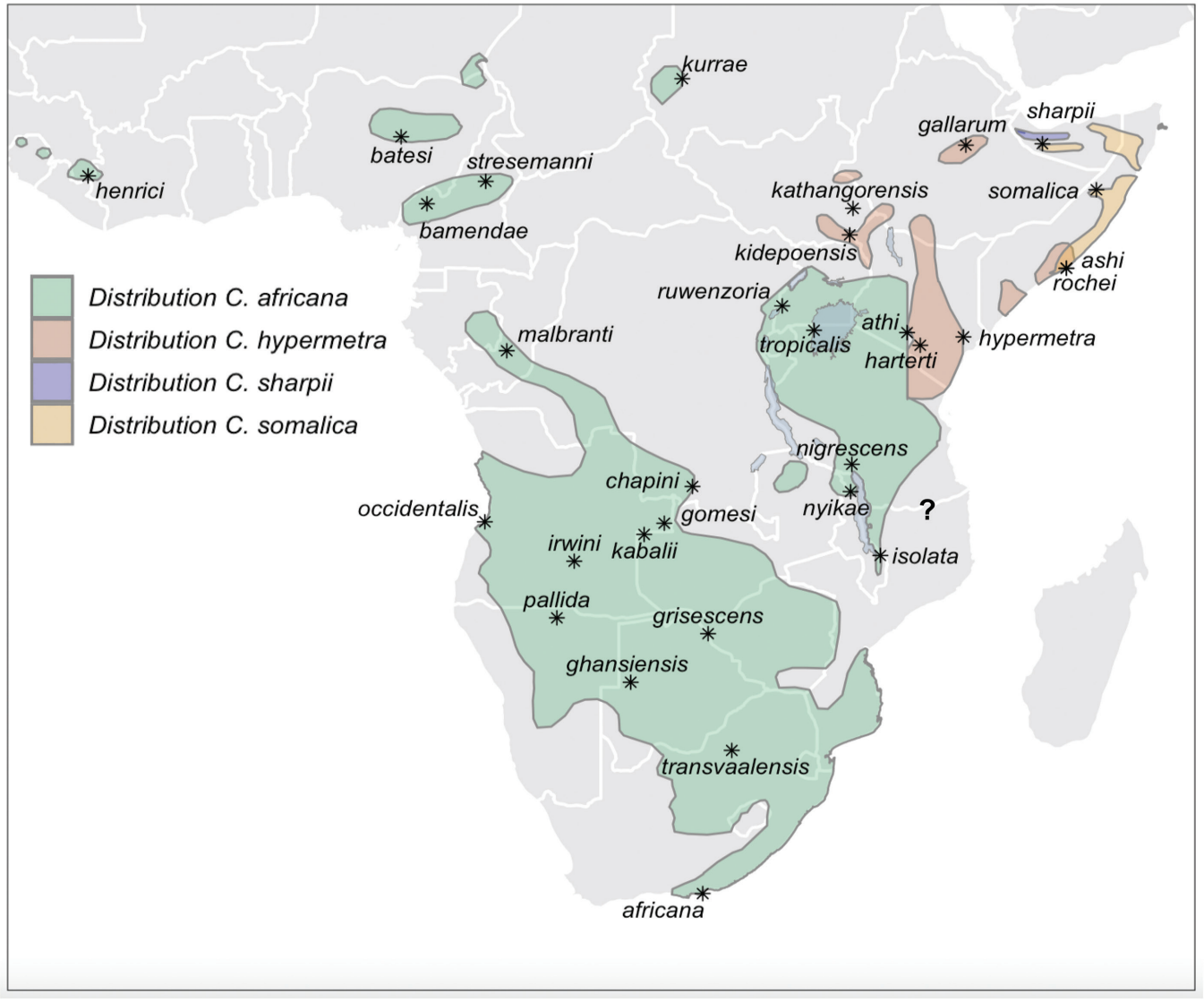

Figure 1. Distribution of the *Corypha africana–C. sharpii–C. hypermetra–C. somalica–C. ashi* complex based on BirdLife International and Handbook of the Birds of the World (2020) and Kennedy and Finch (in prep.), with type localities indicated by asterisks. Note that information on distribution should be regarded with caution, because further research is required for precision and accuracy. The distributions of specific taxa are listed in Table 2, but their boundaries (particularly within and between the two large continuous ranges in southern/south-western Africa and East Africa) are unclear and in need of updating. For example, there have been records further east in southern Tanzania and northern Mozambique (Baker and Baker 2014; Supporting Information, Fig. S1B), indicated on the map by a question mark. Some additional information on distribution can be gleaned from Supporting Information, Figure S1, which shows localities for birds that have been sampled genetically or whose song has been recorded.

DNA sampling and sequencing

From the ingroup, we obtained a total of 38 fresh or toepad samples from 23 taxa classified as *C. africana*, *C. sharpii*, *C. hypermetra*, *C. somalica* and *C. ashi*. This represents 74% of the 31 taxa currently recognised in this complex (Table 2; Supporting Information, Fig. S1; Supplementary Material, Dataset SM1) and includes all the taxa that we considered a priori to belong to distinct groups based on morphology and/or geographical distributions, covering all species suggested to be recognised by Alström *et al.* (2023) based on a multilocus analysis. One other *Corypha* species (*C. fasciolata*), one *Amirafra* [formerly *Mirafra*, see Alström *et al.* 2023; *Amirafra rufocinnamomea* (Salvadori, 1866)] and two species each from the genera *Heteromirafra* [*Heteromirafra ruddi* (Grant, CHB, 1908) and *Heteromirafra archeri* Clarke, 1920], *Galerida* [*Galerida cristata* (Linnaeus, 1758) and *Galerida deva* (Sykes, 1832)] and *Melanocorypha* [*Melanocorypha calandra* (Linnaeus, 1766) and *Melanocorypha yeltoniensis* (Forster, JR, 1768)] were included as outgroups (based on Alström *et al.* 2013a, 2023; Table 2).

For the fresh samples, DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). For detailed descriptions of laboratory procedures for degraded DNA samples from museum toepads, including all procedures, from DNA extraction and USER enzyme (New England Biolabs) treatment to library construction, see Meyer and Kircher (2010) and Irestedt *et al.* (2022). Libraries were whole-genome resequenced on an

Illumina NovaSeq 6000 S4 (Illumina, CA, USA) for 28 of the taxa (35 samples) at a median coverage of 29.7× (interquartile range, 18.1–36.0×; minimum, 6.5×; maximum, 46.6×), and libraries from three of the toepad samples (*C. a. chapini*, *C. a. occidentalis* and *C. h. kidepoensis*) were sequenced at low whole-genome coverage (≤ 1×).

In addition, *Cytb* for four samples of *C. africana athi* was sequenced according to protocols described by Olsson *et al.* (2005).

Phylogenetic multilocus analyses

Sequence assembly

The mitochondrial cytochrome *b* (*Cytb*), nuclear myoglobin intron 2 and partial exons (myo), ornithine decarboxylase introns 6–7 and partial exons (ODC) and recombination activating gene part 1 (RAG1) were harvested from the resequenced genomes in Geneious v.9.1.2 (Biomatters) by mapping first to the target loci in the *Eremophila alpestris* (Linnaeus, 1758) reference genome, then again to the consensus sequence of the assembled reads. For the three toepad samples (*C. a. chapini*, *C. a. occidentalis* and *C. h. kidepoensis*) sequenced at low whole-genome coverage, we only harvested *Cytb* by assembling long contigs of the enriched mitochondrial genome (100–200×) through baiting with lark mitochondrial genes and iterative mapping with MITObim v.1.9.1 (Hahn *et al.* 2013), followed by remapping of reads for visual verification.

Phylogenomic analyses

Single nucleotide polymorphism calling and variant filtering

We called variants relative to the *Eremophila alpestris* reference genome (GCA_009792885.1; Mason *et al.* 2020) for the 35 higher-coverage resequenced samples (27 ingroup samples and eight outgroup taxa), after trimming Illumina adapters with Fastp v.0.20.0 (Chen *et al.* 2018) and evaluating sequence quality with FastQC (Andrews 2010). Read mapping and variant calling were performed using the accelerated wrapper tools in the Sentieon toolkit v.201911 (Freed *et al.* 2017). Reads were mapped using BWA-mem v.0.7.17-r1188 (Li 2013), and PCR duplicates removed using the LocusCollector and dedup modules in Sentieon. We called variable positions using HaplotypeCaller GATK v.4.1 (McKenna *et al.* 2010) and performed joint genotyping of all samples using GenotypeGVCFs (GATK v.4.1) using the Haplotyper and Genotyper modules in Sentieon. Finally, we applied the following common GATK filters to SNPs using the VariantFiltration command in GATK v.4.1: RPRS < −8, QD < 2.0, FS > 60.0, SOR > 3.0, MQ < 40.0 and MQRankSum < −12.5. We also set genotypes with a depth < 2 or depth > 100 and those with a genotype quality < 20 to no call using the SelectVariants command in GATK v.4.1. Filtered SNPs, indels, sites with > 90% missing data, and non-biallelic SNPs were subsequently removed from the callset using BCFtools (Danecek *et al.* 2021).

For phylogenetic analyses, we performed further stringent SNP filtering using VCFtools v.0.1.17 (Danecek *et al.* 2011) to require a minimum genotype quality of 20 (--minGQ 20) and full sample coverage, i.e. none with missing data (--max-missing-count 0), and excluding SNPs mapped to scaffolds ≤ 0.5 Mbp, rendering 1,267,287 SNPs. This '35-sample 1.3M SNP set' was used for analyses with IQ-TREE and served as a starting point for further data reduction for other analyses (see below).

Maximum likelihood tree of 1.3 million SNPs

Using IQ-TREE multicore v.1.6.12 (Nguyen *et al.* 2015), we implemented ModelFinder (Kalyaanamoorthy *et al.* 2017) to identify the best substitution model for the 35-sample 1.3M SNP set, applied ascertainment bias correction (-m MFP+ASC), and calculated node support with 1,000 replicates of ultrafast bootstrap (Hoang *et al.* 2018). In order to obtain a rooted, ultrametric tree, we used the *root* and *chronos* functions in Ape v.5.6-2 (Paradis and Schliep 2019) in R v.4.1.2 (R Core Team 2021). We used discrete evolutionary rates, set the smoothing parameter λ to 10, and calibrated the tree with the approximate 95% highest posterior density ages from Alström *et al.* (2013a) for seven overlapping nodes outside the focal group, connecting: (1) *Galerida* species at 6.0–8.5 Mya; (2) *Melanocorypha* species 4.0–6.0 Mya; (3) *Heteromirafra* species 4.75–8.75 Mya; (4) *Galerida* and *Melanocorypha* 9.75–12.25 Mya; (5) *Galerida*/*Melanocorypha* and all other species (root node) 13.25–17.0 Mya; (6) *Heteromirafra* and *Corypha* 12.0–15.25 Mya; and (7) *Amirafra rufocinnamomea* and *Corypha* 10.25–14.0 Mya.

Species tree inferred from individual nuclear locus trees

ASTRAL-III (Zhang *et al.* 2018) infers a species tree based on individually precomputed trees from discrete loci, such as gene sequences. Applying this to the 35-sample 1.3M SNP set, we mimicked loci by extracting SNPs occurring in non-overlapping 50 kbp windows ('loci') spaced at 1 Mbp intervals, for the ingroup, *C. fasciolata* and *A. rufocinnamomea* (29 samples). This was done for two independent window sets, for which the distance between windows among the sets was 500 kbp (i.e. window set A was extracted at 0.5, 1.5, 2.5 … Mbp of each scaffold, whereas window set B was extracted at 1.0, 2.0, 3.0 … Mbp). In both sets, we retained only windows containing ≥ 10 informative sites, resulting in 626 windows comprising 32,001 SNPs (set A) and 599 windows comprising 28,256 SNPs (set B). For individual windows within the two sets, we obtained maximum likelihood (ML) 'locus trees' using IQ-TREE (Nguyen *et al.* 2015) with the same settings and procedures as described above (see 'Maximum likelihood tree of 1.3 million SNPs'). Contracting branches with very low support improves accuracy of ASTRAL (Zhang *et al.* 2018), hence we used Newick utilities (Junier and Zdobnov 2010) to remove branches with ultrafast bootstrap support ≤ 10%, before running ASTRAL-III v.5.7.8 (Zhang *et al.* 2018). Given that ASTRAL-III estimates terminal branch lengths only for taxa with multiple samples, we only considered the topology and node support (not branch lengths) of the ASTRAL-III species trees.

Coalescence-based species tree analyses

Using VCFtools, we extracted data for the ingroup, *C. fasciolata* and *A. rufocinnamomea* (29 samples), from the 35-sample 1.3M SNP set, which we then thinned by 50 kbp physical distance to avoid linkage. We used vcf2phylip v.2.8 (Ortiz 2019) to convert .vcf files to phylip and binary nexus files comprising 17,467 SNPs.

Snapper v.1.1 is a new, computationally streamlined method to infer a species tree under the multi-species coalescent from SNP data using diffusion models (Stoltz *et al.* 2021), available in

BEAST v.2.7.1 (Bouckaert *et al.* 2019). We imported the binary nexus alignment into Beauti, and we specified Snapper analyses with a beta root prior and log-likelihood correction, applying a Yule model tree prior. Initially, we ran analyses for 50,000–100,000 generations, after which we optimized operators. We ran four replicates of the final analyses for 7,800,000–9,500,000 generations, sampling every 250th generation. We inspected results in Tracer v.1.7.1 (Rambaut *et al.* 2018) for stationarity and convergence between replicate runs. Although the posterior levelled and converged in a relatively efficient manner, certain coalescent rate parameters, and thereby the tree likelihood, continued to display some difference between runs. After discarding 58–92% as burn-in, we combined the runs with LogCombiner, acknowledging low effective sample sizes (ESS) for the parameters mentioned above. We then computed a maximum clade credibility tree with average node heights, using TreeAnnotator, both tools from BEAST v.2.7.1.

SNAPP v.1.5.2 (Bryant *et al.* 2012) also infers a species tree from SNP data under the multi-species coalescent, but unlike Snapper it integrates mathematically over each SNP tree. It is computationally more demanding, which we bypassed in part by including only a single representative per taxon (22 samples and 16,412 SNPs) and by running the ruby script snapp_prep.rb (https://github.com/mmatschiner/snapp_prep; Stange *et al.* 2018) to link theta across branches in order to eliminate the estimation of branch-specific population sizes. We ran linked-theta SNAPP v.1.5.2 analyses in BEAST v.2.6.7 (Bouckaert *et al.* 2019) at four different seeds for 1,500,000 generations, sampling every 500th generation. The output was assessed and processed like Snapper analyses (above) in BEAST v.2.7.1, and for SNAPP analyses we could ensure sufficient ESS (> 200) and convergence between replicates.

Bayesian four-locus analyses

Given that SNPs were not available for all samples, we also assembled a matrix based on *Cytb*, myo, ODC and RAG1 sequences, with the two additional *Corypha* species, *Amirafra* and *Heteromirafra* as outgroups. These sequences were checked and aligned by ClustalW (Thompson *et al.* 1994) implemented in Bioedit v.7 (Hall 1999), together with GenBank sequences from *C. a. transvaalensis* (2), *C. h. gallarum*, *C, apiata* and *Corypha angolensis* (Barboza du Bocage, 1880) (Supporting Information, SM1). The best-fitting models were calculated in jModeltest (Darriba *et al.* 2012): GTR+I+G for *Cytb*, TN93 for myo, HKY+G for ODC and GTR+G for RAG1. This multilocus dataset including 37 ingroup and six outgroup samples was analysed in BEAST v.1.8.4 (Drummond *et al.* 2012). We applied the *Cytb* molecular clock rate of 2.1% divergence/Myr based on Weir and Schluter (2008), setting mean rate to 0.0105 substitutions/site/lineage/Myr and a normally distributed clock prior with SD of 0.001. We set the rate on *Cytb* and allowed the other loci to be estimated. The final analysis was run under an uncorrelated log-normal relaxed clock model (Drummond *et al.* 2006) and a Yule tree prior. Markov chain Monte Carlo (MCMC) chains were run for 200,000,000 generations, sampling every 5,000 generations, on the CIPRES Science Gateway (Miller *et al.* 2010). We used Tracer v.1.7.1 (Rambaut *et al.* 2018) to check the posterior distribution and ESS of the MCMC output (ESS ≥ 200). TreeAnnotator v.2.2.1 (Rambaut and Drummond 2015) was used to summarise trees with mean node height after discarding the first 25% of the trees as burn-in (well after stationarity was reached). The tree and divergence times were displayed in FigTree v.1.7.1 (Rambaut 2016).

Phylogenetic affinity of nuclear SNP alleles in Corypha hypermetra kidepoensis

Corypha h. kidepoensis was sequenced only at low whole-genome coverage, preventing standard variant calling. However, analyses of *Cytb* suggested that *C. h. kidepoensis* was sister to *C. sharpii* rather than to the other *C. hypermetra* subspecies (see Results). To investigate whether this was only a mitochondrial signal, we assessed the nuclear phylogenetic affinity of *kidepoensis* by selecting all sites in the 35-sample 1.3M SNP set where the two *C. hypermetra* samples (*C. h. hypermetra* and *C. h. gallarum*) were fixed for one allele while the two *sharpii* samples were fixed for the alternate allele, tallying which allele the *kidepoensis* reads represented, if present (irrespective of coverage).

Detection of introgression

We extracted data for the ingroup, *C. fasciolata* and *A. rufocinnamomea* from the 35-sample 1.3M SNP set, which resulted in a matrix of 1,226,324 SNPs. We then computed *D* and *f4* statistics in Dsuite v.0.4 r43 (Malinsky *et al.* 2021) for every species trio, with *A. rufocinnamomea* as the outgroup. The F-branch statistic (f_b; Malinsky *et al.* 2018) was subsequently used to characterise excess allele sharing (a sign of potential introgression) among branches of the *Corypha* tree, based on three alternative topologies according to the SNAPP, ASTRAL/IQ-TREE and multilocus trees. The *P*-values were Bonferroni-corrected at $\alpha = 0.01$.

Morphological sampling and analyses

Specimens were examined in the Natural History Museum, Tring, UK (specimen photographs are shown in Supporting Information, Fig. S2), independently and together by P.A., M.S. and P.F.D. For some of the taxa, the sample sizes were very small, in four cases only single adult specimens (of which three were holotypes; Supporting Information, SM2). Comparisons were also hampered by different taxa being collected at different times of the year, hence being at different stage of plumage wear. Series of specimens were compared directly, and differences and similarities were noted, first by P.A., and then the notes were checked by P.F.D. and M.S. Differences were not quantified in any more objective way, such as by using a spectrophotometer. In our opinion, in order for such measurements to be meaningful, they need to be made on live birds, when the different feather tracts can be accessed properly, especially on the upperparts and wings. They also need to be made on birds in comparable plumage, which is often impossible with museum specimens (e.g. some taxa were mainly or exclusively represented by specimens in worn plumage, whereas others were mainly or only represented by fresh specimens).

Measurements of wing length (maximum chord) and tail length were recorded with a ruler to the nearest 0.5 mm. Bill length (to skull), tarsus length, length of the hind claw, length of the outermost primary (P10) and the distance from the tip of P10 to the wing tip were recorded with digital callipers to the nearest 0.1 mm. All measurements were made by M.S., except

some complementary measurement of P10, recorded by P.F.D. A total of 195 specimens (133 males, 51 females and 11 undetermined) were measured and classified to taxon (Supporting Information, SM2), of which 12 had only partial data owing to physical damage.

The dataset of 183 specimens (124 males, 50 females and 9 undetermined) with complete data was explored and analysed in R v.4.1.2 (R Core Team 2021). Given that the sample size was relatively low for several taxa and that most larks exhibit moderate sexual size dimorphism (de Juana *et al.* 2004), we first explored the effect of sex through general linear models for each morphometric variable, with taxon and sex as factors, allowing interaction between the factors. For two measurements (hind claw and P10) there were weakly significant interaction terms between sex and taxon (for three and two taxa, respectively) but no significant main effect of sex. For tail length, there was a weakly significant interaction term between sex and taxon (for three taxa) and a strong effect of sex. For four measurements (wing, tarsus, bill and P10–wing tip) there were no interactions between taxon and sex, while sex as factor had a significant effect. In order to enable analyses of the whole dataset rather than males only, we adjusted the measurements of wing, tarsus, bill, P10–wing tip and tail, for females and unsexed birds, based on the effect of sex determined from rerunning the above models without interaction terms. Given that the distribution showed some degree of skew for most morphometric variables, we $\log_{10}$-transformed the seven variables.

We performed a principal component analysis (PCA) on all seven variables simultaneously using the prcomp function and explored the distribution of eigenvalues using the packages factoextra and corrplot. The PCA was visualized with the function ggplot_pca, using packages AMR (Berends *et al.* 2022) and ggplot2 (Wickham 2009). We also performed linear discriminant analysis (LDA), using various groupings of specimens (main clades or sub-clades, as indicated in phylogenies, and suggested species definitions; see Results and Discussion) as explanatory variables. We fitted different numbers and combinations of morphometrics as dependent variables (scaled to zero mean and unit variance) to the discriminant model, achieving the best results by including all seven. ANOVA and Tukey's post hoc tests were used to identify significant pairwise differences between taxa for specific measurements.

Song sampling and analyses

The typical songs are simple and stereotyped and are usually delivered from a perch, such as a bush, a rock or a mound (e.g. termite hill). More complex songs that include mimicry of other species are occasionally given, sometimes in flight, at least in *africana/transvaalensis*, *tropicalis*, *athi* and *malbranti*. We did not analyse these complex songs, because they are delivered very infrequently, and very few sound recordings are available. We obtained songs of usable quality from 148 different individuals of 18 taxa from xeno-canto (https://xeno-canto.org), Macaulay Library (https://www.macaulaylibrary.org) and the British Library sound archive (https://www.bl.uk/collection-guides/wildlife-and-environmental-sounds), as well as from colleagues and friends (see Acknowledgements) and our own recordings; all previously unpublished recordings have been made available in Macaulay Library (Supporting Information Figs S1B; SM3). Despite our best efforts, for 13 taxa we have been unable to find any sound recordings at all, and for some taxa our sample size is very small. However, we obtained sound recordings from all the primary clades identified in the molecular analyses (see Results), i.e. the proposed species according to Alström *et al.* (2023), except for the single-taxon lineage *sharpii*, which has not been observed since 1975 (Ash and Atkins 2009). The sound terminology is explained in the Supporting Information (Fig. S3). Sonograms were produced in Raven Pro v.1.5 (Bioacoustics Research Program 2017).

For each sound file, three song strophes were measured. In cases where the file contained two or more different song types, three strophes of each unique song type were measured. In total, 207 song strophes with unique patterns were measured; all measurements were made by M.N. and P.A. For each strophe, the following characteristics were measured/noted: duration (in seconds), duration 90% (in seconds), minimum frequency (in hertz), maximum frequency (in hertz), centre frequency (in hertz), bandwidth (in hertz), bandwidth 90% (in hertz), average entropy (in bits), number of syllables, duration of each syllable (in seconds), duration of each gap between syllables (in seconds) and number of elements in each syllable. Bandwidth refers to the difference between maximum and minimum frequency; bandwidth 90% and duration 90% ignore the top and bottom 5%. Centre frequency is the frequency that splits the selected spectrum into two parts of the same energy (Charif *et al.* 2010: p. 171). Average entropy is estimated by taking the average of the entropy in each frame in the selection (Charif *et al.* 2010: p. 170).

For each unique strophe type, the mean values of each of the different measurements were calculated. For the taxa *hypermetra*, *gallarum*, *kidepoensis*, *somalica* and song type 2 of *athi* (see Results), the number of syllables, duration of each syllable (in seconds) and duration of each gap between syllables (in seconds) were not measured because of the different song structure of these taxa (see Results). Instead, the proportion of different element types, the maximum number of times a repeated element/element block was given, the duration of the longest element/syllable, the duration of the shortest element and the duration of the longest gap between elements were measured/noted (otherwise, same as above). All original sound data are in the Supporting Information (SM3).

We used LDA to assess similarity and clustering in song structure in multivariate space between the five major clades uncovered within *C. africana s.l.* (i.e. *africana*, *kabalii*, *athi*, *kurrae* and *nyikae*), excluding the type 2 song of *athi* (see Results). In a separate analysis, we did the same for the songs of *C. somalica*, *C. hypermetra* and the type 2 song of *athi*. We did not include all taxa simultaneously because different variables were measured for each group; some of the metrics for the *somalica*/*hypermetra*/*athi* type 2 vocalisations had no obvious homologues within the broad *C. africana* group. Clinal variation within taxa was assessed by plotting linear discriminant scores of the first two axes against latitude and longitude. In each analysis, we used correlation clustering to identify and remove strongly intercorrelated (at $r > 0.6$) variables. In the LDA analysis, all variables were scaled to zero mean and unit variance owing to their greatly different scales of measurement. Confusion matrices were used to assess the discriminatory power of the models. ANOVA and Tukey's

post hoc tests were used to identify significant pairwise differences between taxa in individual vocal metrics.

RESULTS

Phylogeny

Single nucleotide polymorphism analyses

The SNAPP analyses based on 16,412 SNPs with *Corypha fasciolata* and *Amirafra rufocinnamomea* as outgroup (Fig. 2A) recovers a tree with eight primary clades (I–VIII), with estimated divergence times of 3.0–5.4 Mya and posterior probabilities (PPs) of 1.00 for all nodes except two terminal nodes within clades III and VIII, which are virtually undifferentiated: *C. ashi* and *C. s. rochei* are supported as sisters, diverged 24,000 years ago (24 kya) with a PP of 0.97 and, in turn, diverged from *C. s. somalica* 114 kya. The deepest divergence among *C. a. bamendae*, *C. a. stresemanni* and *C. a. kurrae* is estimated at 12 kya, with the former two recovered as sisters with a PP of 0.61.

Corypha africana s.l. is recovered as paraphyletic, because *C. a. athi* and *C. a. harterti* (clade IV) form a clade with *C. hypermetra* (clade I), *C. sharpii* (clade II) and *C. ashi* and *C. somalica* (clade III). Clades I–IV and V–VII form a large monophyletic group that, in turn, is sister to clade VIII (Fig. 2A). Divergence times within the eight main clades range from 3 kya to 1.4 Mya (Fig. 2A).

The Snapper analyses were based on 17,467 SNPs, which include all sites analysed with SNAPP and an additional 1,000 derived from analysing multiple individuals for some taxa. Replicate runs did not converge as efficiently as SNAPP analyses, and the final tree likelihood varies between −460,520 and −460,511 (SDs of 4–7) and posterior between −460,559 and −460,549 (SDs of 31–37). However, all replicates produce the same backbone tree topology, while clades of more than two closely related taxa have effectively unresolved within-clade topology, with PP < 0.5, and the final tree was therefore produced from the combined replicates (Supporting Information, Fig. S4). The relationship between the eight primary clades is identical to the SNAPP tree (Fig. 2A). The divergence times among the eight primary clades (I–VIII) are estimated to be older than by SNAPP, 3.4–7.7 Mya, while the unresolved within-clade divergence is reconstructed as very young across all main clades (3–50 kya; Supporting Information, Fig. S4).

The ASTRAL tree summarising ML analyses of 32,001 SNPs across 626 unlinked 50 kbp windows (Fig. 2B) differs from the SNAPP tree in placing clade VI as sister to all clades except clade VIII with 100% support (replicated in the separate set of 599 windows comprising 28,256 SNPs; Supporting Information, Fig. S5). The IQ-TREE ML tree, based on 1.3 million concatenated SNPs and including eight outgroup taxa, of which six are closely related species pairs (Supporting Information, Fig. S6), recovers the same topology as the ASTRAL tree, with all nodes receiving 100% bootstrap support except the sister relationship between *kabalii* and *malbranti* (clade VI), which receives 95% bootstrap support. This contrasts to the three SNAPP and ASTRAL trees, in which *kabalii* is sister to *irwini* with PP of 1.00 and 71% support, respectively. The estimated divergence times of clades I–VIII are 2.45–4.9 Mya.

Multilocus analysis

The BEAST analysis based on the mitochondrial *Cytb* and three nuclear loci (5,285 bp) includes three taxa for which only *Cytb* was available: *C. africana chapini*, *C. a. occidentalis* and *C. hypermetra kidepoensis* (Fig. 3). The last taxon (clade II*b*) is sister to *C. sharpii* (clade II*a*) with strong support and rather deep divergence [2.4 Mya, 95% highest posterior distribution (HPD) 1.7–3.3 Mya], whereas the other two belong to the *africana* clade (clade V*a*). In other respects, it resembles the ASTRAL and IQ-TREE trees, except that clade VI is sister to clades V and VII, albeit with extremely weak support (PP of 0.58). Moreover, unlike the other analyses, within clade VI, a sister relationship between *malbranti* and *irwini* is recovered with a PP of 1.00. Divergence times are older than in the SNAPP and IQ-TREE trees, ranging from 4.3 Mya (95% HPD 3.3–5.4 Mya) to 6.8 Mya (95% HPD 5.3–8.6 Mya) between clades I–VIII; clade III and clade VII exhibit particularly deep divergence in comparison to the previous analyses.

Unlike the multilocus study by Alström *et al.* (2023), *C. fasciolata* and *C. apiata* do not form the sister clade to the present clade VIII, but instead form the sister clade to the entire *C. africana*–*C. sharpii*–*C. hypermetra*–*C. somalica*–*C. ashi* complex. The same applies to *C. fasciolata* in the IQ-TREE analysis of SNP data (Supporting Information, Fig. S6).

Nuclear analyses of C. hypermetra kidepoensis *and* C. sharpii

To test the unexpected sister relationship between *C. hypermetra kidepoensis* and *C. sharpii* that was recovered by the mitochondrial *Cytb*, we utilised low-coverage SNPs from *kidepoensis*. For 33,982 SNPs, where *sharpii* and *C. h. hypermetra/gallarum* are fixed for different alleles, *kidepoensis* shares alleles with *sharpii* in 75.1% and with *hypermetra/gallarum* in 24.9%. This pattern is consistent across the genome, with the interquartile range across scaffolds being 72.2–79.5% for alleles shared with *sharpii* and 20.5–27.8% for alleles shared with *hypermetra/gallarum*.

Historical introgression and assessment of tree topology

The Dsuite analyses yield widespread significant f_b values across tree topologies, although most are at low levels (f_b < 0.05; Table 3). The ASTRAL tree topology produces the fewest and the lowest significant f_b values (maximum 0.115), whereas the SNAPP tree topology produces the most and the highest significant f_b values (maximum 0.221; Table 3). With the ASTRAL topology, the three most prominent potential historical introgression events are inferred between the most recent common ancestors of: (1) clade IV–clade V; (2) clade I–clade IV; and (3) clade V (and V*a*)–[*kabalii*, *irwini*] (Supporting Information, Fig. S7A). Applying the SNAPP topology, (1) and (2) above are also recovered, at similar f_b levels as with the ASTRAL topology, but with two additional events at even higher f_b (> 0.20): clade V–clades I–IV and clade V–clade VII (Supporting Information, Fig. S7B). Additionally, although at lower f_b levels, historical introgression is inferred also between clade VII and clades I–IV (Supporting Information, Fig. S7B).

Morphology

The eight most deeply diverged clades in the phylogenetic analyses (I–VIII) are broadly supported by plumage (summary in

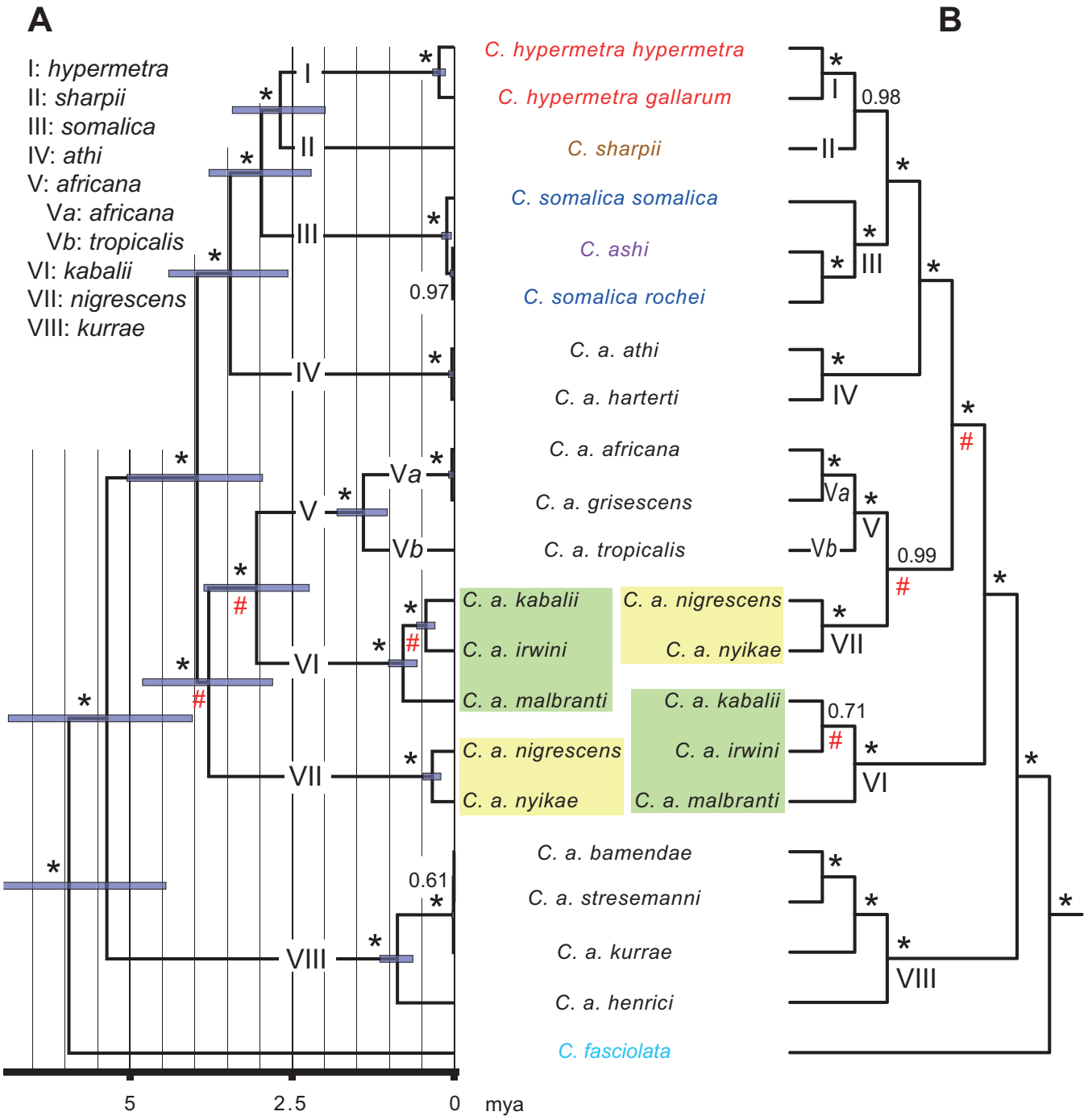

Figure 2. Phylogenies of the *Corypha africana–C. sharpii–C. hypermetra–C. somalica–C. ashi* complex, with *Corypha fasciolata* as the outgroup (the more distant outgroup species *Amirafra rufocinnamomea* has been removed) based on the multi-species coalescent, produced within a Bayesian framework in SNAPP on 16,412 single nucleotide polymorphisms (SNPs) and single individuals per taxon (A) and with ASTRAL, summarising taxon quartets in maximum likelihood (ML) analyses of 32,000 SNPs across 626 unlinked windows for all samples (B). Node support values are posterior probabilities (PPs) for SNAPP and local PPs for ASTRAL, with an asterisk indicating a PP of 1.00. Clades that are discussed in the text are labelled I–VIII and with the name of the taxon that has priority in the respective clade. The names of *C. africana s.l.* are in black font; the other species are in other colours. Topological incongruence between these two trees, the Snapper species tree (Supporting Information, Fig. S3), the ML IQ-TREE tree (Supporting Information, Fig. S5) and/or the Bayesian multilocus tree (Fig. 3) are indicated by a red '#'.

Table 4; a full description of this table is provided in Supporting Information, SM4, with expanded descriptions in Supporting Information, SM5; a photographic overview is given in Supporting Information, Fig. S2; photographs of plumage details are in Supporting Information, Fig. S8). However, no individual trait is unique to a single clade, because traits are partly

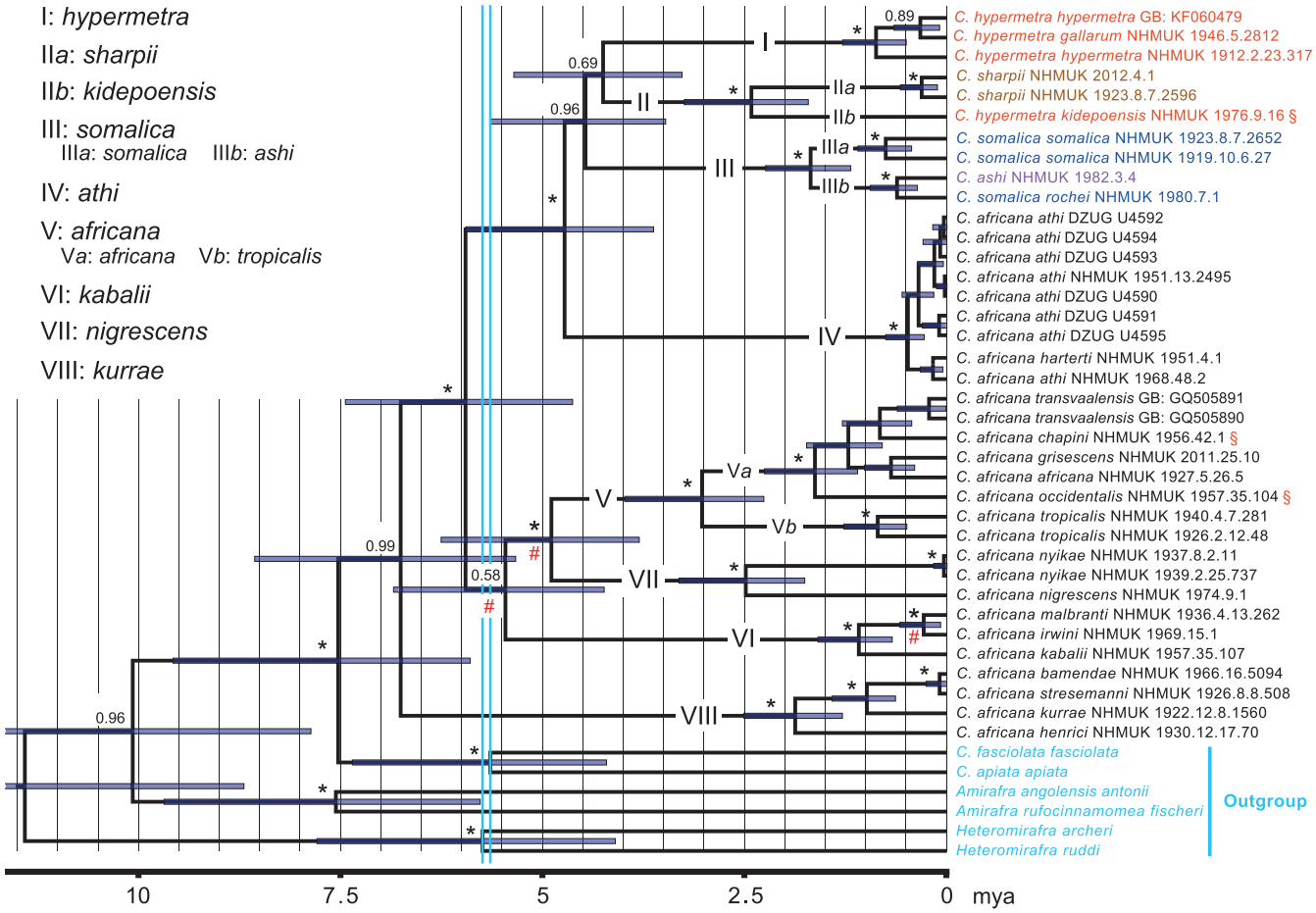

Figure 3. Multilocus tree for the *Corypha africana–C. sharpii–C. hypermetra–C. somalica–C. ashi* complex based on Bayesian analysis (BEAST) of concatenated mitochondrial and nuclear loci (5,285 bp). Posterior probabilities (PPs) are indicated at the nodes, with an asterisk indicating a PP of 1.00. Clades that are discussed in the text are labelled I–VIII and with the name of the taxon that has priority in the respective clade. The names of *C. africana s.l.* are in black font; the other species are in other colours (the six outgroup species are all in pale blue). The three taxa that are not present in any other analyses are indicated by a red '§'. Topological incongruence compared with other phylogenetic trees is indicated by a red '#'. The blue lines indicate the divergence times of two of the outgroup species pairs.

Table 3. Summary of the *f* branch statistic (f_b) values calculated with Dsuite, based on different tree topologies (only significant f_b values are included).

	Tree topology		
	SNAPP	**ASTRAL**	**Multilocus**
Nf_b	135	116	126
Σf_b	5.57	3.58	4.77
$\bar{x} f_b$	0.041	0.031	0.038
Maximum f_b	0.221	0.115	0.168
$Nf_b > 0.05$	37	23	31
$Nf_b > 0.10$	12	3	11
$Nf_b > 0.15$	6	0	1
$Nf_b > 0.20$	2	0	0

or completely overlapping between different clades. Moreover, in some clades there is considerable variation among (and also within) taxa. Nevertheless, based on a combination of characters, all eight main phylogenetic clades are broadly supported by plumage. In general, differences are most pronounced in fresh plumage, diminishing with wear and bleaching.

Morphometrics (Supporting Information, SM2) also broadly support the eight main clades, summarised for each measurement per clade in Supporting Information, Figure S9 and per taxon/taxon group in Supporting Information, Table S1. The PCA based on seven morphometric measurements partitions 49.8% of the variation along the first principal component (PC1), 24.8% along PC2 (cumulative 74.6%) and 11.2% along PC3 (cumulative 85.8%; Supporting Information, Fig. S10) and separates clades I (the '*hypermetra* clade'), II (the '*sharpii* clade'), VII (the '*nigrescens* clade') and VIII (the '*kurrae* clade'; Fig. 4). Clade VI (the '*kabalii* clade') is moderately distinct, as is clade III*a* (*C. s. somalica*), whereas clade III*b* (*C. s. rochei* and *C. ashi*) overlaps with clades IV (the '*athi* clade') and V (the '*africana* clade'; in particular V*b*, the '*tropicalis* clade'; Fig. 4). The LDA yields high classification accuracy: 90.1% and 90.7% when grouped by numbered clades and subclades (cf. Figs 2, 3) and by proposed species delimitation (see Discussion), respectively (Table 5).

Table 4. Summarised categorization of plumage differences among taxa in the *Corypha africana–C. sharpii–C. hypermetra–C. somalica–C. ashi* complex.

Clade	Sub-clade	Taxon	Crown	Mantle/scapulars (fresh)	Base colour of breast to undertail-coverts	Breast streaking	Streaking on flanks	Streaking on undertail-coverts	Bases to primaries + outer secondaries	Tertials + inner greater coverts (fresh)	R5–R6
V[a]	Va	*africana/transvaalensis* (*N* > 10)	1	1	1	1	1	1	1	1	1
		grisescens (*N* = 7)	1	1	1	1	1	1	1	1	1
		ghansiensis (*N* = 2)	1x	1x	1x	1	1	1	1	2*	(1)
		pallida (*N* = 4)	1x	1y	1	1	1	1	1	–	1
		chapini	1	1	1	1	1	1	1	1	1
		occidentalis	1	1	1	1	1	1	1	1	1
		gomesi (*N* = 1)	1	1z	1	1	1	1	1	2x*	–
	Vb	*tropicalis* (*N* > 10)	1	1	1	1	1	1	1	3	1
		ruwenzoria (*N* > 10)	1	1	1	1	1	1	1	3	1
VII		*nigrescens* (*N* = 2)	4	2	1	1	2	2	2	4	2
		nyikae (*N* = 6)	5	3	1	1	2	2	2	5	2
VI[b]		*kabalii* (*N* = 6)	6	4	1	2	1	1	1	6	1
		irwini (*N* = 2)	7	5	1	3	1	1	1	6	1
		malbranti (*N* = 1)	8	–	1	3	1	1	2	–	(1)
VIII[c]		*kurrae* (*N* = 4)	9	6	2	1	1	1/2	2	7	1
		batesi (*N* = 1)	9	6	2	1	1	1	2	7	1
		henrici (*N* = 2)	9	6x	2	1	1	2	2	7x	1
		bamendae (*N* = 4)	9	6x	2	1	1	2	2	7x	1
		stresemanni (*N* = 3)	10	6y	2	4	1	2	2	7y	1
IV		*athi* (*N* > 10)	4	1v	1	1	1	1	1	8	1
		harterti (*N* = 3)	11	1w	1	1	1	1	1	8	1
I		*hypermetra* (*N* > 10)	4	7	1	1	1	2	1	9	1
		gallarum (*N* = 9)	1	7	1	1	1	2	1	9	1
II	IIb[d]	*kidepoensis* (*N* = 6)	1	7x	1	1x	1	1/2	1	9	1
		kathangorensis (*N* = 1)	1y	7x	1	1	1	1	1	9	1
	IIa	*sharpii* (*N* = 6)	13	8	3	5	1	1	1	10	3
III	IIIa	*somalica* (*N* = 8)	13	8x	3	6	2	2	2	11	4
	IIIb	*rochei* (*N* = 1)	13	–	3x	7	2	2	2	–	4
		ashi (*N* = 5)	14	9	3x	7	2	2	2	12	4

Here, each state (variant) of a particular plumage character is represented by a number (where, e.g. 1x is a variation of 1) and colour; white boxes represent unique states, and '–' indicates that no reasonably fresh specimens have been examined or that the character has not been studied. For a full description of each character in each taxon, see the table in the Supporting Information (SM4); see also the expanded descriptions in the Supporting Information (SM5).

[a]*ghansiensis*, *pallida* and *gomesi* have not been sequenced.

[b]The taxa in this clade are characterized further by their less distinct supercilia and less distinctly streaked nape in comparison to the other taxa.

[c]*batesi* has not been sequenced.

[d]*kathangorensis* has not been sequenced.

*The two examined specimens of *ghansiensis* and single *gomesi* have very fresh plumage, and it is possible that with wear, they turn out not to be distinct from other taxa in clade V (category 1 or 3).

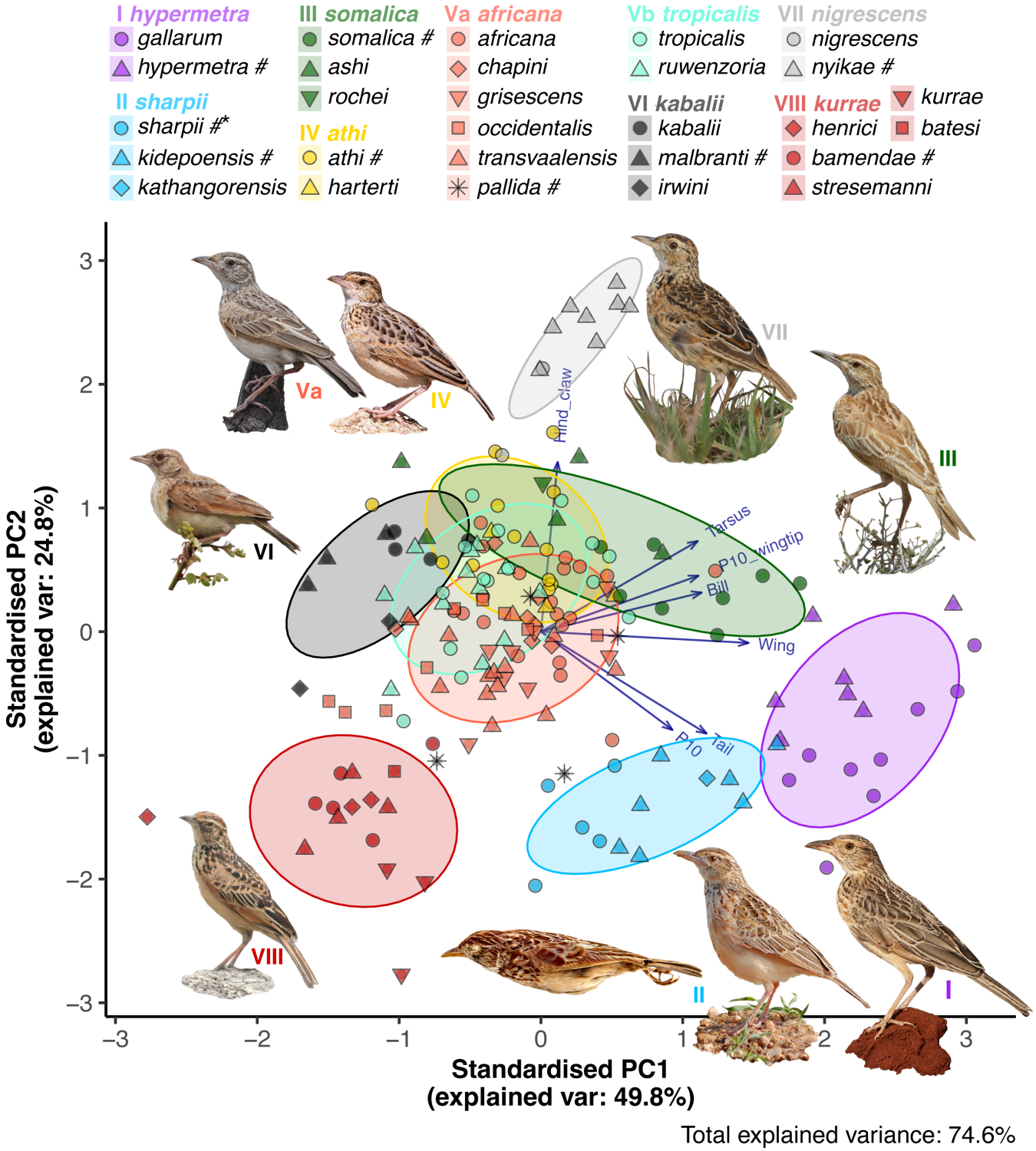

Figure 4. Principal component analysis (PCA) based on seven morphometric variables: the lengths of wing, bill, tarsus, tail, hind claw and outermost primary, and the distance from the outermost primary to the wing tip. Ellipses indicate the 68% confidence interval (±1 SD) for each clade. Loadings for each trait, drawn with maps and their relative contributions across all principal components (PCs), can be viewed in the Supporting Information (Fig. S9). Photographs are by Nik Borrow (II, *kidepoensis* Kidepo National Park, Uganda, May; III, *somalica* Tuuyo Plains, Somaliland, September; VII, *nyikae* Nyika National Park, Malawi, December), Maans Booysen, Birding Weto, CC BY-SA 4.0 https://creativecommons.org/licenses/by-sa/4.0, via Wikimedia Commons (VI, *malbranti* Longa, Angola), Paul F. Donald (Va, *pallida*, Caprivi Strip, Namibia, November), Emmanuel Barde Elisha (VIII, *bamendae* Mambilla Plateau, Nigeria), Stratton Hatfield (I, *hypermetra* Tsavo East National Park, Kenya, March), Peter Steward (IV, *athi* Nairobi National Park, Kenya, March) and The Trustees of The Natural History Museum, London (II, *sharpii* NHMUK, identity and collection information lacking). The taxa with photographs are also indicated by '#', and '*' for *sharpii*.

Table 5. Confusion matrix of linear discriminant analysis based on seven morphometric variables: the lengths of wing, bill, tarsus, tail, hind claw and outermost primary, and the distance from the outermost primary to the wing tip.

A	**Prediction (clade)**										**B**	**Prediction (species)**								
	I	**IIa**	**IIb**	**IIIb**	**IIIa**	**IV**	**Va**	**Vb**	**VI**	**VII**	**VIII**			*afr*	*ath*	*hyp*	*kab*	*kid*	*kur*	*nig*	*sha*	*som*
I: *hypermetra*	15	0	1	0	0	0	0	0	0	0	0	*africana*		83	3	0	2	0	1	0	1	0
IIa: *sharpii*	0	4	1	0	0	0	0	0	0	0	0	*athi*		5	10	0	0	0	0	0	0	0
IIb: *kidepoensis*	1	0	6	0	0	0	0	0	0	0	0	*hypermetra*		0	0	16	0	0	0	0	0	0
IIIb: *rochei*	0	0	0	4	1	0	1	0	0	0	0	*kabalii*		0	0	0	9	0	0	0	0	0
IIIa: *somalica*	0	0	0	0	8	0	0	0	0	0	0	*kidepoensis*		0	0	1	0	6	0	0	0	0
IV: *athi*	0	0	0	0	0	14	0	1	0	0	0	*kurrae*		1	0	0	0	0	15	0	0	0
Va: *africana*	0	2	0	0	0	1	56	2	2	0	0	*nigrescens*		0	1	0	0	0	0	9	0	0
Vb: *tropicalis*	0	0	0	0	0	2	0	24	0	0	1	*sharpii*		0	0	0	0	1	0	0	4	0
VI: *kabalii*	0	0	0	0	0	0	0	0	9	0	0	*somalica*		1	0	0	0	0	0	0	0	13
VII: *nigrescens*	0	0	0	0	0	1	0	0	0	9	0											
VIII: *kurrae*	0	0	0	0	0	0	1	0	0	0	15											

Clade

Species (suggested delimitation)

Predictions are based on: (**A**) clades, as labelled in Figures 2–5; and (**B**) suggested species delimitations, with overall accuracy of 90–91%.

Songs

The eight primary clades (I–VIII) are broadly supported by song. Although we were not able to obtain sound recordings from all taxa, we have analysed songs of representatives of all primary clades, which correspond to the species proposed by Alström *et al.* (2023) (although *sharpii* is missing, and its placement together with *kidepoensis* in clade II rests on *Cytb* and SNPs). In all taxa, there is considerable individual variation in the typical song, which complicates comparisons among taxa. However, in *C. africana s.l.* the individual repertoires are fairly small, with a maximum of six different phrases (strophes) recorded in a single male (Supporting Information, SM3). The repertoires are larger in the *hypermetra* clade (≤ 18 phrases) and in *somalica* (≤ 10 phrases; Supporting Information, SM3).

The songs can be separated broadly into two types. The first of these is given by *athi* (clade IV); *africana*, *transvaalensis*, *grisescens*, *pallida*, *tropicalis* and *ruwenzoria* (clade V); *kabalii* and *malbranti* (clade VI); *nyikae* and *nigrescens* (clade VII); and *bamendae* (clade VIII; Fig. 5: strophes 1–40). These songs are built up of short, di- or trisyllabic whistles, which are interspersed by silent pauses of a few seconds [the durations of pauses depend on the level of excitement (P.A., personal observation) and have therefore not been measured]. The main differences among the groups (clades) concern the duration of the phrases, the maximum frequency, the presence or absence of frequency modulations (as indicated by average entropy), the relative durations of syllables, and the duration of the gaps between syllables (Fig. 6; expanded descriptions in Supporting Information, SM5; original data and mean values in Supporting Information, SM3). There are significant differences between taxa in all these variables (see below).

The second song type is given by *hypermetra* and *gallarum* (clade I); *kidepoensis* (clade II*b*); *somalica* (clade III); and *athi* (clade IV; Fig. 5: strophes 41–57). This differs from the first type by, for example, having longer strophes (mean ± SD: 1.3 ± 0.72 s, vs. 0.68 ± 0.16 s in the first category) with a higher number of elements (mean ± SD: 7.4 ± 3.7, vs. 3.1 ± 0.98 in the first category) and lower centre frequency (mean ± SD: 3.1 ± 0.37 kHz, vs. 3.9 ± 0.43 kHz). In *athi*, the two song types are either given separately for long periods or alternated between, and type 1 is more common than type 2. The songs of *hypermetra* and *gallarum* can be divided into two sub-types: a common type (Fig. 5: strophes 50–56) and a less common song type (Fig. 5: strophe 57), which is often alternated with the commoner type. However, the classification into these two categories is not always clear-cut. The second one is more stereotyped than the first and consists of more monotonous series of whistles (mean ± SD: 74 ± 30.4% different elements in a strophe, vs. 94 ± 11.1% in the commoner song type) of lower bandwidth than in the commoner type (mean ± SD: 2.0 ± 0.5 kHz, vs. 2.9 ± 0.44 kHz in the commoner type). Dowsett-Lemaire and Dowsett (1978) reported mimicry of 20 different bird species in songs of *C. hypermetra*, which were usually 'interspersed with notes of the proper song'. We have not studied the use of mimicry in this species, but Françoise Dowsett-Lemaire (personal communication) found evidence of limited mimicry in three of the eight xeno-canto recordings that we analysed. The song of *somalica* can also be divided into two types, but this classification is even less clear than in *hypermetra*. The main differences among

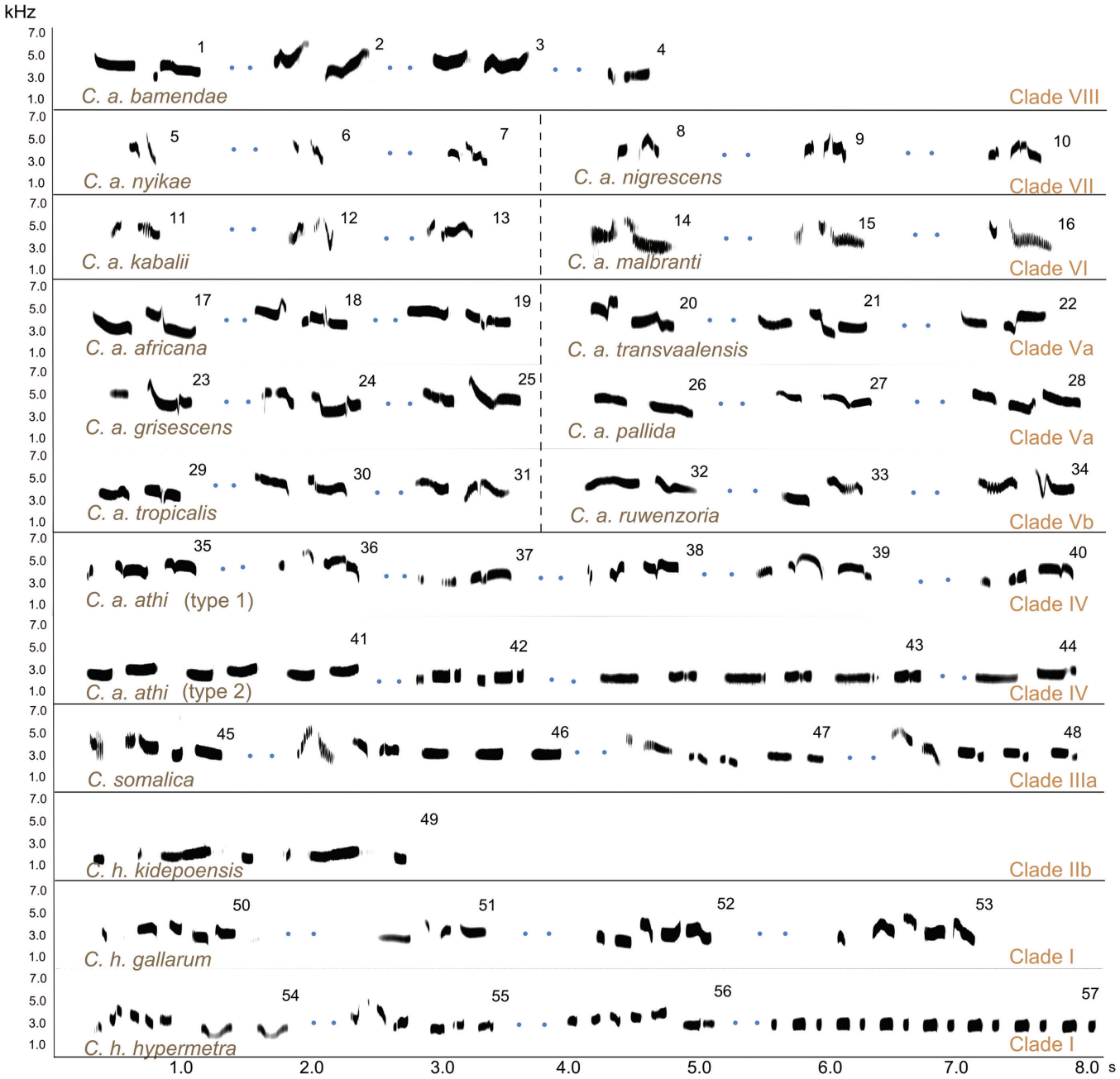

Figure 5. Sonograms of songs of taxa treated as subspecies of *Corypha africana*, *C. hypermetra* and *C. somalica*. The strophes are numbered consecutively for precise reference in the text. Recordings: (1) Foumban-Tibati, Cameroon, BL 022A-W1CDR0000054 BD19 (Claude Chappuis); (2) Mambilla Plateau, Nigeria, ML274689 (Emmanuel Elisha); (3) Mambilla Plateau, Nigeria, ML274691 (Emmanuel Elisha); (4) Mambilla Plateau, Nigeria (Emmanuel Elisha); (5) Nyika National Park, Malawi, ML278996 (Peter Boesman); (6) Nyika National Park, Malawi, ML279004 (Peter Boesman); (7) Nyika National Park, Malawi, ML279005 (Peter Boesman); (8) Kipengere Plateau, Tanzania, ML562576011 (David Moyer); (9, 10) Kipengere Plateau, Tanzania, ML562576201 (David Moyer); (11) Minyanya plain, north-west Zambia, ML279942 = XC509007 (Peter Boesman); (12) Minyanya plain, north-west Zambia, ML279943 = XC509008 (Peter Boesman); (13) Minyanya plain, north-west Zambia, ML279957 = XC509010 (Peter Boesman); (14) Lékoni, Gabon, ML274685 (Nik Borrow); (15) south-east Gabon, Chappuis CD8 BD91 (P. Christy); (16) Lékoni, Gabon (Michael Mills); (17) Polokwane, Limpopo, South Africa, ML274677 (Per Alström); (18) Sandveld, South Africa, XC28667 (Patrik Åberg); (19) Free State, South Africa, XC536937 (Dawie de Swardt); (20) Polokwane, South Africa, ML274681 (Per Alström); (21) North West Province, South Africa, XC516080 (Frank Lambert); (22) Polokwane, Limpopo, South Africa, XC516146 (Frank Lambert); (23) Lusaka, Zambia, BL WS3371 C15 (Robert Stjernstedt); (24) Lusaka, Zambia, BL WS3371 C15 (Robert Stjernstedt); (25) Lusaka, Zambia, BL WS3371 C10 (Robert Stjernstedt); (26) Etosha National Park, Namibia, ML42945 (Linda Macaulay); (27) Etosha National Park, Namibia, ML61179 (Linda Macaulay); (28) Namutoni, Namibia, XC58667 (Charles Hesse); (29) Maasai Mara, Kenya, ML274675 (Stratton Hatfield); (30) Maasai Mara, Kenya, ML274674 (Stratton Hatfield); (31) Mara Triangle, Kenya, XC200107 (Rory Nefdt); (32) Entebbe, Uganda, ML8031 (Myles E. W. North); (33) Uganda, ML23453 (G. Stuart Keith); (34) Semliki Flats, Uganda, ML23454 (G. Stuart Keith); (35) Lake Nakuru, Rift valley, Kenya, ML8032 (Myles E. W. North); (36) Nairobi, Kenya, ML8034 (Myles E. W. North); (37) Nairobi, Kenya, ML274661 (Per Alström); (38) Lake Nakuru, Rift

the groups (clades) concerns, for example, the duration of the strophes, the speed of delivery (number of elements per second) and the presence of rasping/ringing notes (Fig. 6; Supporting Information, SM3).

The two main song types are so different from each other that they were analysed separately. There is no significant clinal variation within taxa based on correlation between song variables and latitude or longitude when *P* is corrected for multiple testing. For the six clades singing the first song type, the LDA incorporates eight variables after reduction of multicollinearity using correlation clustering: duration, minimum frequency, maximum frequency, bandwidth 90%, average entropy, duration of first syllable, duration of second syllable, and duration of first gap between syllables. The first two linear axes account for 97.3% of the variance between taxa. A plot of these two axes (Fig. 7A) and the confusion matrix (Table 6A) suggest that the *athi*, *kabalii*, *nigrescens* and *africana* clades are all separated, with 83.7–96.4% of songs correctly allocated by the model, whereas the *kurrae* clade is poorly separated, with most songs being assigned to the *africana* clade. The overall classification accuracy is 83.9%.

For the analysis of *hypermetra*, *somalica*, *kidepoensis* and the type 2 song of *athi*, the LDA incorporates nine variables after reducing multicollinearity using correlation clustering: duration, minimum frequency, maximum frequency, centre frequency, song elements per second, proportion of different elements, duration of longest element/syllable, duration of shortest element/syllable, and duration of longest gap between syllables. A plot of the first two LDA axes (Fig. 7B) and the confusion matrix (Table 6B) suggest that discrimination between taxa is good, with an overall classification accuracy of 88.9%.

Singing behaviour

The taxa *africana*, *transvaalensis*, *pallida*, *grisescens*, *tropicalis* and *ruwenzoria* (clade V), *kabalii* (clade VI) and *bamendae* (clade VIII) are known frequently to raise their crown feathers when singing, displaying the rufous bases to the rear crown feathers. In contrast, the taxa *athi*, *hypermetra* and *gallarum*, which are relatively well known, apparently do not raise their crown feathers (P.A., P.F.D., D.E., pers. obs. in the field and study of photographs on, e.g., https://www.macaulaylibrary.org/). None of the photographs that we have seen of the following, more poorly known, taxa have displayed raised crown feathers: *kidepoensis*, *harterti*, *malbranti*, *somalica*, *nyikae* and *nigrescens* (for *nigrescens*, also based on field observations by David Moyer, pers. comm.). However, the fact that none of the few photographs we have seen of these taxa (two to seven individuals per species, usually not singing) depict birds with their crown feathers raised does not mean that they might not do this. We have no information on whether any of the other taxa might raise their crown feathers.

Representatives for all clades except III (and II*a*, for which there are no recordings) have been observed to quickly flutter ('clap') their wings in connection with singing, although this is apparently rare in *athi*. The wing clapping can be associated with a small hop, and we have observed *hypermetra* to ascend ~1 m while clapping its wings and then to descend on raised, slowly beating (but not clapping) wings. Although our sample of recordings of wing clapping is small, there appear to be pronounced differences among some taxa with regard to the number of claps, the timing of the wing clapping in relation to the song, and the strength of the clapping (Table 7).

Geographical distributions

For most taxa, the detailed geographical distributions are extremely poorly known, and the maps in Figures 1 and Supporting Information, Figure S1 should be considered to show only approximate distributions of the previously recognised species [based on BirdLife International and Handbook of the Birds of the World (2020) and Kennedy and Finch (in prep.), with type localities indicated in Fig. 1, sometimes slightly outside the currently recognised distributions of the species]. The maps in Supporting Information, Figure S1, with our genetic and vocal samples indicated, give a slightly better representation of the ranges (again, with some samples outside the presently recognised distributions of the species). All taxa in the complex are mainly or entirely para- or allopatric, with the exception of the poorly known *somalica–sharpii* and *hypermetra–ashi/rochei*, which appear to occur in sympatry (Ash and Miskell 1998, Ash and Atkins 2009). It also seems likely that *kabalii/irwini* (and possibly *malbranti*) and *chapini/gomesi/occidentalis* are sympatric, at least locally, although there is no firm evidence of this. The taxa *hypermetra* and *athi* are generally considered to be rather widely overlapping (e.g. Ryan *et al.* 2020). However, critical examination of the evidence indicates that there is probably at the most a very marginal overlap (Kennedy and Finch in prep.), and they have been observed in sympatry at only a handful of localities (Tyler Davis, Brian Finch, Stratton Hatfield, Washington Wachira, in litt.). The distribution of *nigrescens/nyikae* is localised and restricted to high plateaus bordering northernmost Lake Malawi (Fig. 1; Supporting Information, Fig. S1), with elevational segregation from other taxa. The main distribution of *transvaalensis* is in southern Africa, but it has also been suggested to occur in southern Tanzania (Table 2; Britton 1980, Baker and Baker 2014). It is unclear whether the gap of ≥ 300 km between southern African and East African populations within the *C. africana s.l.* complex represents true absence or a lack of data.

Valley, Kenya, ML8032 (Myles E. W. North); (39) Nairobi, Kenya, ML8035 (Myles E. W. North); (40) Nairobi, Kenya, ML274662 (Per Alström); (41) Nairobi, Kenya (Michael Mills); (42) Nairobi, Kenya, ML 274669 (Per Alström); (43) Along route Nairobi–Tsavo, Kenya, ML274699 (Per Alström); (44) Nairobi NP, Kenya, ML 274668 (Per Alström); (45) west of War Idaad, Somaliland (Michael Mills); (46) west of Ceel Afweyn, Somaliland (Michael Mills); (47) west of War Idaad, Somaliland (Michael Mills); (48) west of Ceel Afweyn, Somaliland (Michael Mills); (49) Kidepo valley national park, Uganda, BL WA 2017/006/039/13 (Michael Mills); (50) Awash national park, Ethiopia (Michael Mills); (51) Aledeghe Wildlife Reserve, Ethiopia, XC267268 (Andrew Spencer); (52) Aledeghi Wildlife Reserve, Afar, Ethiopia (Andrew Spencer); (53) Awash National Park, Ethiopia, ML100296 (Linda Macaulay); (54) Tsavo West, Kenya, ML274697 (Per Alström); (55) Taita Taveta, Kenya, XC178954 (Richard Dunn); (56) Tsavo East, Kenya (Michael Mills); (57) Tsavo West, Kenya, ML274697 (Per Alström). Abbreviations: BL, British Library; ML, Macaulay Library; XC, xeno-canto.

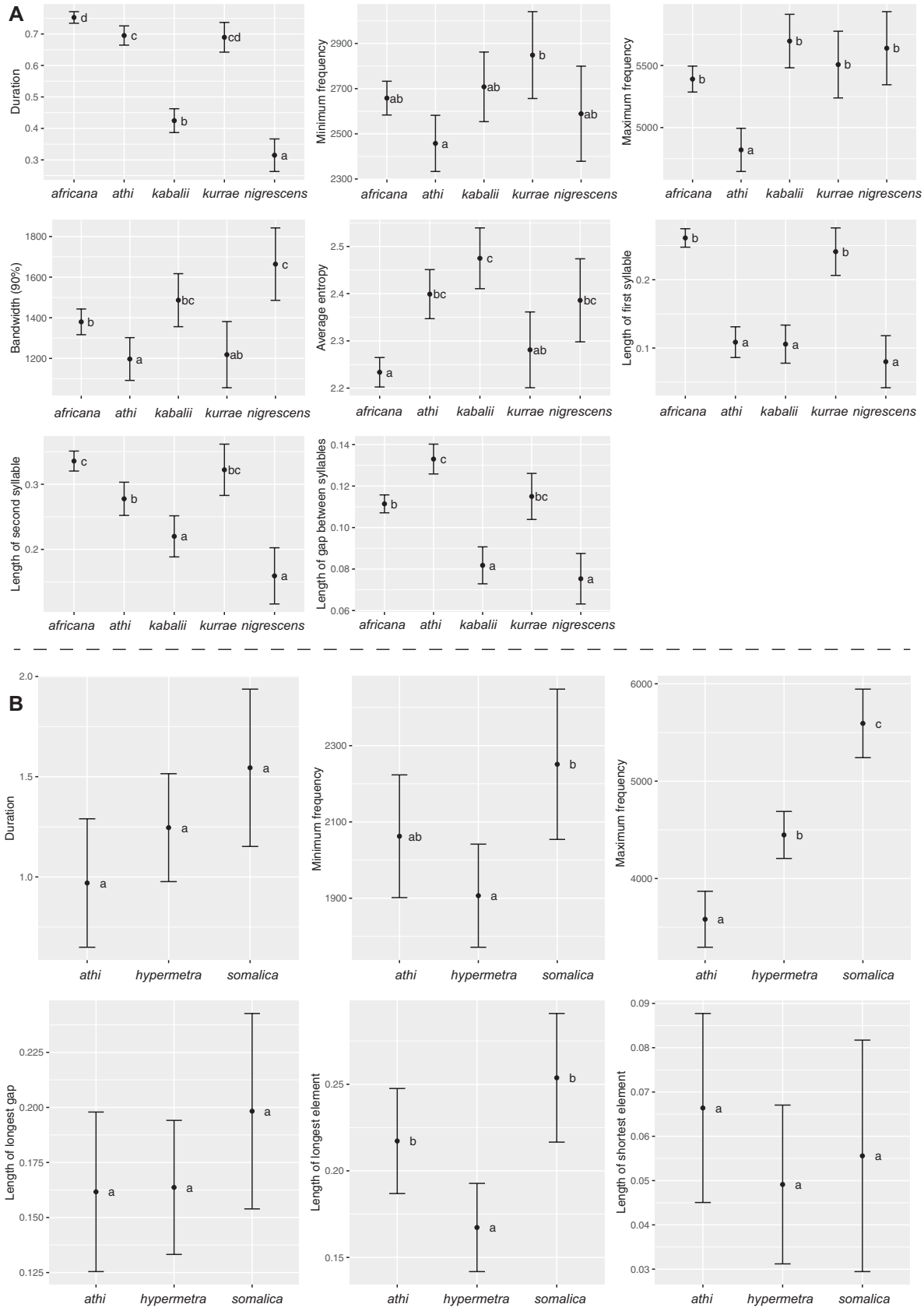

Figure 6. A, mean and SD for eight song variables for the five primary clades in the *Corypha africana* complex (first song type, corresponding to Fig. 5: strophes 1–40). B, mean and SD for six song variables for type 2 song of *C. a. athi*, and song of *Corypha hypermetra* and *C. somalica* (corresponding to Fig. 5: strophes 41–57).

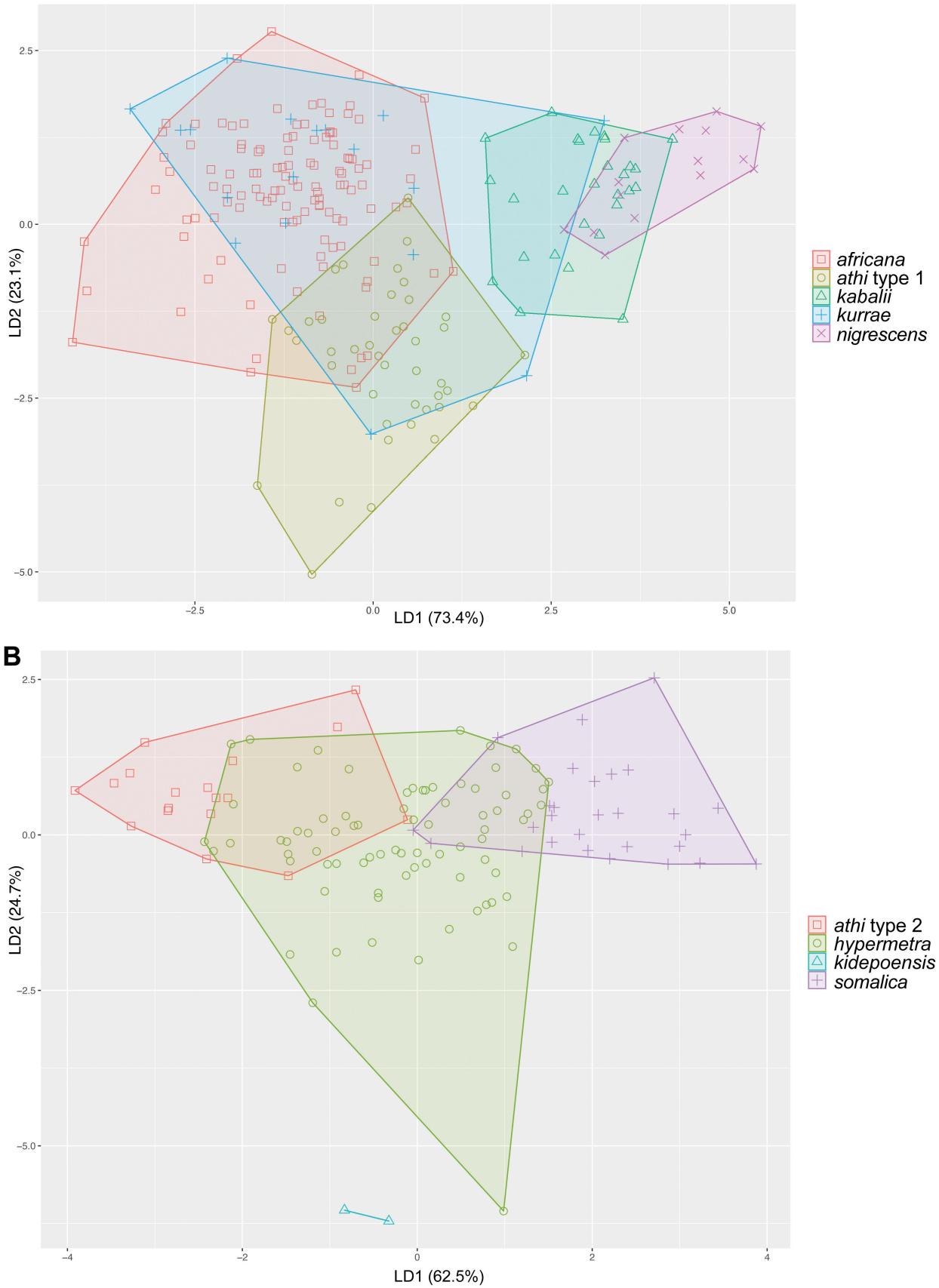

Figure 7. Plots of the first two components of a linear discriminant analysis of songs among the six primary clades in the *Corypha africana* complex, excluding the type 2 song of *C. a. athi* (A) and songs among *Corypha hypermetra* (*C. h. hypermetra* and *C. h. gallarum*), *C. h. kidepoensis*, *C. somalica* and the 'type 2' song of *C. a. athi* (B).

Table 6. Confusion matrix of linear discriminant analysis based on songs of: (**A**) the five primary clades of *C. africana s.l.* (excluding *athi* type 2 song); and (**B**) *C. hypermetra*, *C. hypermetra kidepoensis*, *C. somalica* and *C. africana athi* type 2 song.

A	**Prediction (clade)**						**Classification accuracy (%)**
Clade (actual)	**V**	**IV**	**VI**	**VIII**	**VII**		
V: *africana* (N = 119/68)	111	8	0	0	0		93.3
IV: *athi* type 1 (N = 43/27)	7	36	0	0	0		83.7
VI: *kabalii* (N = 28/9)	0	0	27	0	1		96.4
VIII: *kurrae* (N = 18/9)	14	2	0	1	1		5.6
VII: *nigrescens* (N = 15/9)	0	0	3	0	12		80.0

B	**Prediction (clade)**				**Classification accuracy (%)**
Clade (actual)	**IV**	**I**	**IIb**	**IIIa**	
IV: *athi* type 2 (N = 18/12)	15	3	0	0	83.3
I: *hypermetra* (N = 78/17)	4	70	1	3	89.7
IIb: *kidepoensis* (N = 2/1)	0	0	2	0	100.0
IIIa: *somalica* (N = 26/7)	0	3	0	25	89.3

The N values are given as number of unique strophes/number of individuals.

DISCUSSION

Many of the numerous taxa in the widespread *Corypha* complex have been poorly characterised, with unknown phylogenetic relationships. Based on our integrative analyses, we propose that nine species should be recognised and that one of the traditional species (*C. ashi*) should not be recognised at species rank (Fig. 8; Table 2). Most of these species are supported congruently by multiple lines of evidence, most importantly deep divergence and paraphyly of the currently recognised species, complemented by differences in morphology, vocalisations and other behaviours (Fig. 8). The same nine species were suggested to be recognised by Alström *et al.* (2023) based on multilocus analyses of all except one of the species in the family Alaudidae.

Phylogeny

All analyses, of ≤ 1.3 million genome-wide SNPs and of a small number of nuclear introns and *Cytb*, reconstruct the same eight primary clades (I–VIII). The relationships among these clades are concordant except for the position of clade VI. In the SNAPP (Fig. 2A) and Snapper (Supporting Information, Fig. S4) analyses, clade VI is recovered as sister to clade V, whereas ASTRAL (Fig. 2B) and IQ-TREE (Supporting Information, Fig. S6) recover clade VI as sister to all other clades except clade VIII. These rather different positions are equally well supported in the different analyses. Finally, the multilocus analysis places clade VI as sister to clade V + VII, with low support (Fig. 3). Furthermore, clade VI is the only clade whose internal taxon relationships are discordant among analyses, with all three possible combinations strongly supported in alternative analyses (Fig. 2).

The above incongruences could be the result of incomplete lineage sorting, hemiplasy (Avise and Robinson 2008) or ancient introgression (Edwards *et al.* 2016, Taylor and Larson 2019, Rancilhac *et al* 2021, Zhang *et al.* 2021). Although all species tree methods do account for incomplete lineage sorting, gene flow (introgression) violates their assumptions. Analyses of excess allele sharing (Table 3; Supporting Information, Fig. S7) consistently recover two plausible cases of introgressive hybridisation, between the most recent crommon ancestors of clade V (*africana*, *tropicalis*, etc.) and clade IV (*athi*, *harterti*) and between the latter and the most recent common ancestor of clade I (*hypermetra*, *gallarum*), with both pairs currently distributed parapatrically in the vicinity of the Eastern Rift Valley. Within clade VI, no gene flow could be detected between *malbranti* and either *kabalii* or *irwini* based on the SNAPP and ASTRAL topology (*malbranti*,(*kabalii*,*irwini*)), while introgression is inferred between the two latter taxa and clade V (and/or V*a*). This agrees with the present-day distributions, where *malbranti* is isolated in eastern Gabon, south central Congo and south-western Democratic Republic of Congo, while *kabalii* and *irwini* are almost surrounded by, or perhaps even sympatric with, taxa that are confirmed (*grisescens*, *chapini* and *occidentalis*) or presumed (*pallida* and *gomesi*) to be members of clade V*a* (Fig. 1). In the SNAPP topology, the different position of clade VI breaks the monophyly of a large clade composed of clades I–V + VII. Consequently, based on this topology, we recover an excess of allele sharing between the ancestral branch of clade V and those of both clade VII and clades I–IV. However, this pattern could be obtained in the absence of introgression along those branches,

Table 7. Wing fluttering ('clapping') in the taxa in which this behaviour has been recorded.

Taxa	Mean; median; range; number of bursts; number of individuals	When in relation to song; perceived strength
Clade Va		
africana & transvaalensis	3.8; 3; 2–7; 54; 10	Just before song; faint
pallida	6; 6; 5, 7; 2; 2	Just before song; faint
grisescens	3.7; 3.5; 2–5; 6; 2	Just before song; faint
Clade Vb		
tropicalis & ruwenzoria	7; 7; 6–8; 8; 4	Just before song; rather strong
Clade VI		
kabalii	19.1; 18; 16–29; 15; 5	During song (starting just before, ending just after song); strong
Clade VII		
nyikae	10; 10; 9–11; 4; 2	Before song; rather strong
nigrescens	10; 11; 7–12; 3; 1	Before song; rather strong
Clade IV		
athi	12 strokes; 1 burst	Before song
Clade I		
hypermetra	7.7; 8; 7–8; 1; 1	During song (beginning immediately before or just after start of song); rather strong
gallarum	14.5; 14.5; 10, 19; 2; 2	Mainly before song (ending just after beginning of song in one individual, ending after c. 1/3 of song in second individual; rather strong
Clade IIb		
kidepoensis	18 strokes; 1 burst	Separate from song; strong

if clade V was more closely related to clades I–IV + VII than to clade VI. Thus, the sister relationship between clades V and VI in the SNAPP tree could be explained by a single introgression event along ancestral branches of both clades, as supported by the analyses based on the ASTRAL topology.

Inferring introgression patterns in the presence of phylogenetic uncertainty is a notoriously complicated problem (Beckman *et al.* 2018, Pease 2018), because interpreting an excess of allele sharing relies on assumptions on the species relationships (Patterson *et al.* 2012). In our case, the conflicting position of clade VI results in variations in the number of introgression events and their strength depending on the considered topology. However, we argue that the ASTRAL topology is more parsimonious, in that it requires fewer introgression events, which are also more consistent with the present distribution of taxa (although this argument is only tentative, because ranges can change quickly and drastically). Although introgressive hybridisation appears to be much more common than was formerly thought, genome-wide signals of introgression are expected to remain scarce, because gene flow will cease after speciation in most cases, and introgressed alleles at low frequency are likely to disappear quickly. Furthermore, hybridisation is a non-random process that requires a specific geographical setting (i.e. that both populations are in contact), meaning that not all possible introgression events are equally likely. Thus, we argue that the Dsuite results might be used as an argument in favour of the ASTRAL topology over the SNAPP tree. However, it should be emphasized that regardless of the topology, some introgression events are consistently supported, meaning that a strictly bifurcating tree cannot represent the evolutionary history of the *Corypha* complex accurately. Why ASTRAL would infer more accurate relationships than SNAPP remains unclear, because both methods rely on the same assumptions. However, it is possible that introgressed regions have been highly fragmented by recombination, in which case the relatively large windows used for the ASTRAL analysis would be less likely to support the introgressed locus tree, in comparison to the single SNPs in the SNAPP analysis.

Taxonomy

Phylogenetics

The fact that our phylogenetic analyses strongly suggest that *C. hypermetra*, *C. sharpii*, *C. somalica* and *C. ashi* are all nested within *C. africana s.l.* calls for the following taxonomic change: either (i) treatment of all five species as a single species; or (ii) splitting of *C. africana* into multiple species. The divergence times support species status for the eight most deeply diverged clades (I–VIII). These are separated by ~4.3–6.8 Myr in the multilocus analysis and by 2.7–5.9 Myr (SNAPP) or 2.45–4.9 Myr (IQ-TREE, calibrated with outgroup nodes) in the SNP analyses. Importantly, the second youngest of these divergence times (4.5 Mya in the multilocus, 3.1 Mya in the SNAPP or 2.8 Mya in the IQ-TREE) concern taxa that are sympatric: *hypermetra* (I) vs. *ashi* and *rochei* (III*b*), and *sharpii* (II) vs. *somalica* (III*a*) (Ash and Miskell 1998, Ash and Atkins 2009). Moreover, in the IQ-TREE tree, the divergence times between all primary clades except I–III vary from ~0.5 Myr younger to 1.0 Myr older in comparison to that between *Melanocorypha calandra* and *Melanocorypha yeltoniensis* (Supporting Information, Fig. S6). These two Palaearctic *Melanocorypha* larks are widely sympatric and strikingly different in appearance (Cramp 1988, Shirihai and Svensson 2018). In the multilocus analysis, which also includes several outgroup species (but not the two *Melanocorypha* species), clades V–VIII vary from 0.8–0.9 Myr younger to 0.9–1.0 Myr older than the two species pairs *Corypha apiata–C. fasciolata* and *Heteromirafra archeri–H. ruddi*.

Further support comes from the recent multilocus analysis of all except one of the currently recognised species of larks

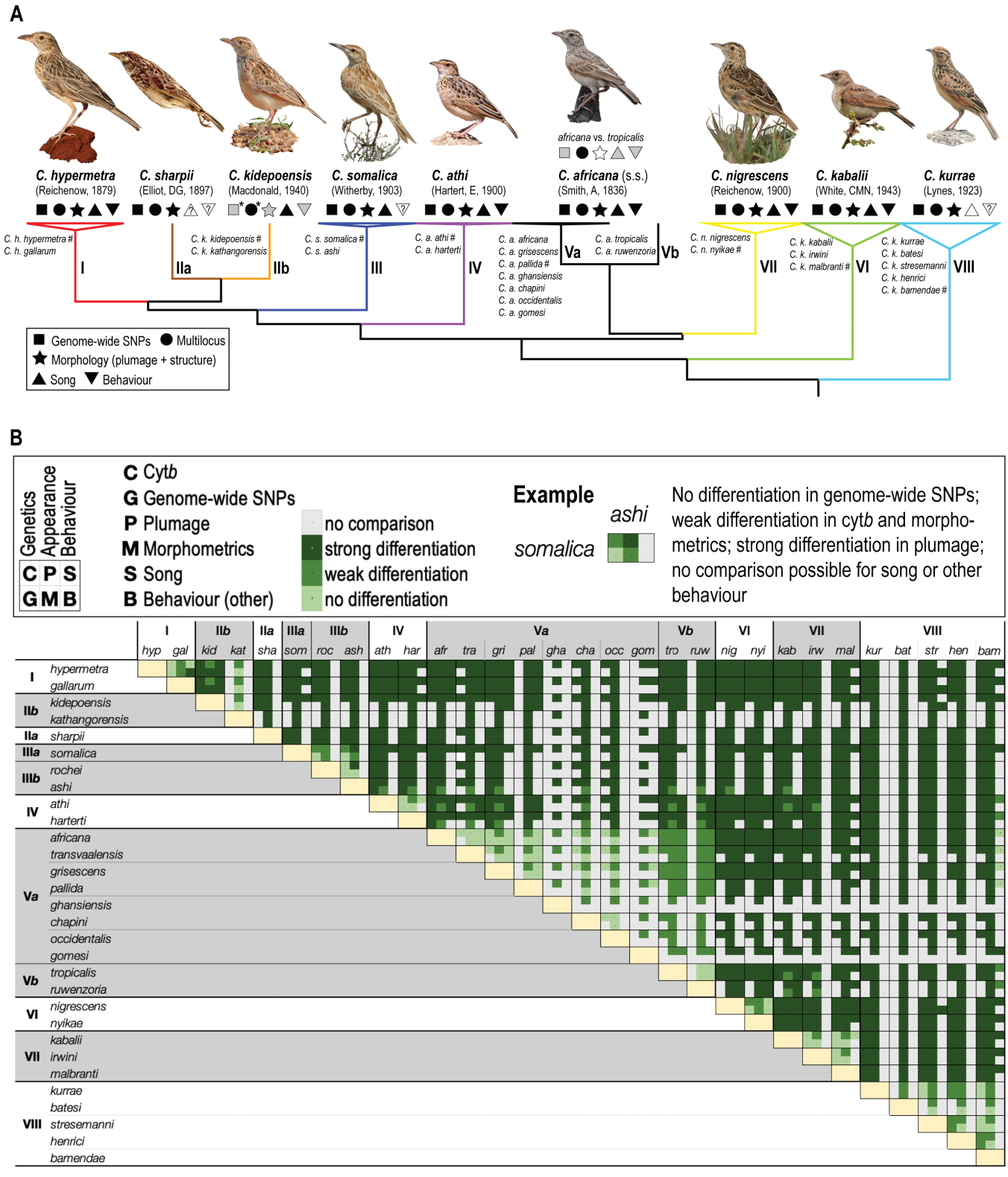

Figure 8. Summary of evidence. A, tree with support for different clades based on different datasets, labelled with the proposed new taxonomy. Symbols are explained at the top: black symbols = strong support; grey symbols = weak support; unfilled symbols = no support; white triangle with question mark = lack of data. *Restricted datasets were analysed for *Corypha kidepoensis*: only *Cytb* in the multilocus analyses, and a separate single nucleotide polymorphism (SNP)-based analysis not included in the principal component analysis (see main text). #Taxon featured in photograph. B, pairwise differentiation between all taxa, based on mitochondrial DNA, genome-wide SNPs, plumage, morphometrics, song and other behaviour. The full key and an example are at the top of the panel. For photograph credits, see Figure 4.

by Alström *et al.* (2023). In that study, the divergence times between the clades corresponding to the present clades I–VIII were estimated at 3.5–5.7 Mya, with the divergence time between clades II*a* (*sharpii*) and II*b* (*kidepoensis*) 2.0 Mya (95% HPD 1.0–3.0 Mya). In comparison, the same study found four sympatric and six parapatric or narrowly sympatric lark species pairs that were more recently diverged than the eight primary *Corypha* clades, and two parapatric or narrowly sympatric species pairs that were more recently diverged than *sharpii* (II*a*) and *kidepoensis* (II*b*). For example, the divergence times between the species in the sympatric, but ecologically partly or largely segregated, and morphologically strikingly different sister pairs, Tibetan lark *Melanocorypha maxima* Blyth, 1867–Mongolian lark *Melanocorypha mongolica* (Pallas, 1776) (Alström 2020a, 2020b), calandra lark *Melanocorypha calandra*–black lark *Melanocorypha yeltoniensis* (Cramp 1988, Shirihai and Svensson 2018) and Bengal bushlark *Plocealauda* (previously *Mirafra*) *assamica* (Horsfield, 1840)–Indian bushlark *Plocealauda erythroptera* (Blyth, 1845) (Alström 1998, Rasmussen and Anderton 2012), were estimated at 2.7–2.9 Mya.

Corypha a. tropicalis (V*b*) exhibits fairly deep divergence from the *africana* clade (V*a*) in all analyses: 1.4 Myr in the SNAPP analysis and 1.6 Myr in the IQ-TREE analysis based on SNPs, and 3.0 Myr in the multilocus analysis. In the multilocus analysis by Alström *et al.* (2023), *tropicalis* is separated from *africana* + *transvaalensis* by 2.5 Myr, i.e. almost as anciently diverged as the sympatric *P. assamica*–*P. erythroptera*. Based on this alone, *tropicalis* might qualify for species status (but see below). Fairly deep divergences are also found within clades VI, VII and VIII, although not across all analyses.

In contrast to the above-mentioned deep divergences, *C. ashi* is essentially undifferentiated from *C. somalica rochei* and only weakly diverged from *C. s. somalica* in the SNP trees (Fig. 2A; Supporting Information, Figs S4–S6). The divergences among these three taxa are considerably deeper in the multilocus tree (Fig. 3). However, in the latter analysis the divergence between the two *somalica* samples is deeper than between *ashi* and *rochei*, indicating that these results should be viewed with caution. These results suggest that *C. ashi* should be treated as conspecific with *C. somalica* (see further under 'Nomenclature'). Another surprising finding is the strongly supported sister relationship between *C. hypermetra kidepoensis* (clade II*b*) and *C. sharpii* (clade II*a*), based on a single *Cytb* sequence from the former (Fig. 3). Although the low whole-genome resequencing coverage of *kidepoensis* does not allow for the high-confidence SNP calling required for traditional nuclear phylogenetic analyses, the sister relationship to *C. sharpii* is corroborated by analyses showing that *kidepoensis* shares more nuclear alleles with *sharpii* than with *hypermetra/gallarum*. Accordingly, based on these results, *kidepoensis* should either be treated as conspecific with *sharpii* or as a distinct species (see further under 'Morphology and vocalisations' and 'Taxonomic and nomenclatural recommendations').

Morphology and vocalisations

Based on a combination of plumage and structure, all taxa form groups that conform largely with the eight primary clades (I–VIII). However, *sharpii* (clade II*a*) does not clearly group with *kidepoensis* and *kathangorensis* (clade II*b*) based on plumage, although morphometrics support this grouping. The similarity in plumage between *kidepoensis/kathangorensis* and *hypermetra/gallarum* might be plesiomorphic, in which case the distinctness of *sharpii* is attributable to this taxon having diverged more in plumage than the others in clade I + II. This is a common pattern within Alaudidae, as suggested by Alström *et al.* (2013a). Some of the taxa currently treated as subspecies of *C. africana* are so strikingly divergent from each other in plumage and structure that it is hard to understand on what basis they have been considered conspecific. For example, the large, proportionately short-tailed and long-clawed *nyikae* is strikingly different in plumage and all other aspects from the small, relatively long-tailed and short-clawed *kabalii*, *malbranti* and *irwini* (Table 4; Supporting Information, Fig. S2; SM2, SM4 and SM5). The last three taxa are also markedly different from each other in plumage (Table 4; Supporting Information, Figs S2, S8; SM2, SM4 and SM5). In fact, the morphological differences across clades I–VIII are considerably greater than the differences between some Palaearctic taxa that are treated as separate species, such as the Eurasian skylark *Alauda arvensis* Linnaeus, 1758 vs. Oriental skylark *Alauda gulgula* Franklin, 1831 (Alström 2020c, Campbell *et al.* 2020), the Asian short-toed lark *Alaudala cheleensis* vs. Turkestan short-toed lark *Alauda heinei* (von Homeyer, 1873) vs. Mediterranean short-toed lark *Alaudala rufescens* (Vieillot, 1819) (Alström *et al.* 2021), and the greater short-toed lark *Calandrella brachydactyla* vs. Mongolian short-toed lark *Calandrella dukhunensis* (Sykes, 1832) vs. Hume's lark *Calandrella acutirostris* Hume, 1873 (Stervander *et al.* 2016, 2020, Alström and Gombobaatar 2021, Alström and Sundev 2021, Shirihai *et al.* 2023). Nevertheless, plumage colours and patterns are known to be correlated with the substrate on which larks occur (Donald *et al.* 2017); therefore, differences in plumage on their own might not be considered very important.

Six of the eight primary clades (I, III–VII) are supported by the song analyses. This also applies to clade II*b*, while no recordings are available for the single taxon in clade II*a* (*sharpii*). The songs of *hypermetra/gallarum* (clade I), *somalica* (clade III) and *kidepoensis* (clade II*b*) are so markedly different in structure from the songs of the *C. africana s.l.* taxa (excluding the type 2 song of *athi*, see below) that we measured partly different variables in these two groups and analysed them separately. In the LDA, *hypermetra/gallarum*, *somalica* and *kidepoensis* are well separated, with 89–100% of the songs correctly classified, and it seems likely that all could be identified by ear and examination of sonograms. However, our sample of *somalica* is rather small (26 unique strophes from seven individuals), and we have only two unique strophes from one *kidepoensis* and no *ashi* or *rochei*. Although we have sound recordings of only 13 of 23 taxa within *C. africana s.l.*, and small series of several of these, the songs support four (IV–VII) of the five primary clades of *C. africana s.l.*, with 83–96% of the songs correctly classified in the LDA. In the *nigrescens* clade (VII), 3 of 15 strophes were mis-classified as the *kabalii* clade (VI) based on the measured variables. However, these two are easily separable by examination of sonograms and by ear: *nigrescens* and *nyikae* in the former clade lack the distinct frequency modulations/thin rattles of *kabalii* and *malbranti* in the latter clade. The only primary clade that is poorly supported by song is the *kurrae* clade (VIII), with only 5.6% of the recordings correctly classified, while 78% are classified as the *africana*

clade (V). With respect to *tropicalis/ruwenzoria* (clade V*b*), only 66% of the songs are correctly classified, with 23% classified as the *africana* clade (V*a*). The taxon *athi* (clade IV) is further supported by having a second song type, which is more similar to some of the songs of *hypermetra*, *gallarum*, *kidepoensis* and *somalica* than to any of the other *C. africana s.l.* taxa (or to the usual, type 1 song of *athi*).

Given that songs are generally considered to be of great importance for mate choice in passerine birds, hence for reproductive isolation between species (Price 2008, Päckert 2018), they are highly relevant in taxonomic studies (reviewed by Alström and Ranft 2003). It might therefore seem surprising that the song of *bamendae* in the *kurrae* clade (VIII; no other taxa in this clade studied) exhibits particularly poor differentiation from several other taxa, especially the *africana* clade (V), although clade VIII is consistently reconstructed as sister to all other clades, with a divergence time 4.9–6.8 Mya. However, several sympatric lark species have extremely similar songs, e.g. the calandra lark *Melanocorypha calandra* and the bimaculated lark *M. bimaculata* (Ménétries, 1832), and the crested lark *Galerida cristata*) and Thekla's lark *G. theklae* Brehm, AE, 1857 (Shirihai and Svensson 2018). It is evident that song evolution, like divergence in other traits, does not proceed at equal rates across lineages. Lack of differentiation in songs should not be taken as evidence of lack of lineage divergence, whereas divergence in songs could indeed be evidence of lineage separation (cf. Sangster 2018). Moreover, given that nearly all taxa are para-/allopatrically distributed, with the *kurrae* clade being particularly isolated from the other clades (Fig. 1), there has been no opportunity for selection for reinforcement of differences (nor reinforcement of discrimination ability; Butlin 1987, Hudson and Price 2014).

Both the number of wing claps and the perceived strength and timing of the clapping in relation to the song differ among some of the taxa, further supporting the distinctness of some clades. However, field observations are lacking for 18 of the 31 taxa, hence more work is needed. The behaviour to raise the crown feathers to form a small crest while singing also differs among taxa, although information is lacking for more than half of the taxa.

Nomenclature

The taxon *irwini* is usually not recognised (e.g. Clements *et al.* 2022, Gill *et al.* 2023); however, the two specimens we have examined (NHMUK 1969.15.1 and NHMUK 1969.15.2, from Cuito-Canavale, Cuando-Cobango, Angola, i.e. an area from where the majority of the specimens mentioned in the type description originated; da Rosa Pinto 1968) are highly distinctive (Supporting Information, Fig. S2), and the genomic divergence from *kabalii* and *malbranti* is marked (Fig. 2A; Supporting Information, Fig. S6), and we therefore recognise this taxon. Nevertheless, it would be desirable to examine additional material of this and geographically nearby taxa.

Corypha ashi, which is unanimously recognised as a species in the literature (e.g. del Hoyo and Collar 2016, Clements *et al.* 2022, Gill *et al.* 2023), is here shown genetically to be poorly differentiated or undifferentiated from *C. somalica rochei*. Although the available specimens of *ashi* and *rochei* look different, the small sample sizes and different plumages (five fresh specimens of the former and one worn specimen of the latter) preclude judgement of these plumage differences. These samples were collected at the same locality (Uarsciek, also known as Warsheikh, ~60 km north-east of Mogadishu in southern coastal Somalia) and, with the exception of one *rochei* specimen that we have not examined (see Roche 1966), represent the entire collection of the taxa. We suggest that *rochei* should be treated as a synonym of *ashi* (both described in the same paper by Colston 1982).

Data-deficient taxa

For eight taxa, we have no sequence data (Table 2). For only one of these, *isolata*, we lack data altogether. This taxon was synonymised with *transvaalensis* by Peters (1960) without further comments, although it is widely separated geographically from *transvaalensis* from southern Africa. It was recognised by Dickinson and Christidis (2014) with the comment 'For recognition see Dean et al. 1992'. However, there are no comments supporting this taxon in the referenced work (cited here as Dean and Keith 1992), only a four-word description that could match any taxon. The type description (Clancey 1956), which was based on 'two freshly moulted examples', does not indicate whether it really is distinctive or not, although it states explicitly that it 'cannot be confused' with *nyikae*. We tentatively recognise it.

For two of the other taxa for which we lack molecular data (*pallida* and *ruwenzoria*) we have morphological, vocal and other behavioural data, which indicate strongly that they are closely related to the taxa in the *africana* clade.

With respect to *ghansiensis*, we have studied only two topotypical specimens, which are in very fresh plumage, unlike the specimens of the adjacent *grisescens* and *pallida* that we have examined. This taxon was synonymised with *pallida* by Clancey (1956) without further comments, based on a total of four specimens of *pallida* including *ghansiensis*. However, it was recognised by, for example, Dean and Keith (1992) and Dean (2005). We tentatively recognise *ghansiensis*, because the two specimens we have examined are distinctive, although they certainly appear to be closely related to the taxa in the *africana* clade.

For *batesi* and *kathangorensis* we have had access only to single specimens (holotypes of both; possibly the only specimens existing of these taxa), but their relationships as part of clade VIII and as sister to *kidepoensis*, respectively, are highly probable based on morphology, and are also plausible based on geographical distributions (Fig. 1).

For *gomesi* we have had access only to photographs of a single specimen in the American Museum of Natural History, which resembles the geographically adjacent *chapini* to the extent that it might prove to be a junior synonym. However, while describing this taxon White (1944) noted that '150 miles to the north occurs the totally different *C. a. chapini* Grant & Mackworth-Praed, of which I have now obtained further material from the north of Mwinilunga and northwards, close to its type locality, in the Katanga'. Moreover, Robert J. Dowsett (in litt.) considers it to be valid based on 'good series examined'.

Taxonomic and nomenclatural recommendations

We propose that nine species should be recognised: *C. kurrae*, *C. kabalii*, *C. nigrescens*, *C. africana* (*s.s.*), *C. athi*, *C. somalica*, *C. hypermetra*, *C. sharpii* and *C. kidepoensis* (Fig. 8; Table 2). Most of these are supported congruently by multiple lines of evidence, most importantly deep divergence (Fig. 2) or/and paraphyly of the currently recognised species (Fig. 1), complemented by differences in morphology (Fig. 4; Table 5; Supporting Information, Figs S2, S8; SM4 and SM5), vocalisations (Figs 5–7; Supporting Information, SM5) and other behaviours (Table 7). Among parapatric/sympatric taxa of different species, *C. athi* and *C. a. tropicalis* are the most similar in appearance (detailed in Supporting Information, SM5), especially in worn plumage, although their songs differ (Figs 5–7; Supporting Information, SM5). However, all nine proposed species are phenotypically diagnosable (summarised in Fig. 8; photographs in Supporting Information, Fig. S2; details in Supporting Information, SM5). Accordingly, they qualify as separate species under the general lineage concept (de Queiroz 1998, 1999, 2007) and the phylogenetic species concept (Cracraft 1989, 1997) and are appropriately treated as species also under the biological species concept (Mayr 1942, 1963, 1969). *Corypha ashi* is 'downgraded' to subspecies rank: *C. somalica ashi* (Fig. 8; Table 2).

The recommended taxonomy is supported further by the near-complete phylogenetic analysis of the family Alaudidae, including representatives for all the species proposed here (Alström *et al.* 2023). Our proposed species *C. kidepoensis* (with subspecies *C. k. kidepoensis* and *C. k. kathangorensis*) might be somewhat controversial. Our analyses conclude that *kidepoensis* is sister to *sharpii* and not to *hypermetra/gallarum*, with which it is usually considered to be conspecific. Although the phylogenetic analysis producing this relationship is based on the mitochondrial *Cytb*, it is corroborated by analyses of nearly 34,000 SNPs. In addition, our single sound recording (two different strophe types) of *kidepoensis* is markedly different from the song of *hypermetra/gallarum* (100% correct classification in the LDA), and there are also plumage and structural differences from *hypermetra/gallarum*. The taxon *kathangorensis* is known from only a single specimen, the holotype, which is very similar in appearance to *kidepoensis*, and we therefore tentatively include that as a subspecies of *C. kidepoensis*, while noting that larger sample sizes might even suggest that *kidepoensis* and *kathangorensis* are better treated as synonymous. Both *kidepoensis* and *kathangorensis* are separated geographically from *hypermetra* by the Eastern Rift Valley, while *gallarum* occurs also within the Eastern Rift Valley, yet still some ≥ 600 km away. As an alternative to species status for *C. kidepoensis*, one could treat it as conspecific with *C. sharpii*, based on the sister relationship between these (clades II*b* and II*a*, respectively). However, the plumages of *sharpii* and *kidepoensis/kathangorensis* are markedly different in multiple aspects (Supporting Information, Fig. S2; SM5), and the divergence between *kidepoensis* and *sharpii* is fairly deep (estimated at 2.4 Mya based on *Cytb*), although shallower than the divergence between the narrow-sense *africana* clade (V*a*) and *tropicalis* (clade V*b*) (see next paragraph).

The taxon *tropicalis* (clade V*b*; including *ruwenzoria*, which we propose is better treated as a junior synonym of *tropicalis*, by priority; cf. Table 2) clearly represents a distinct lineage (i.e. a species *sensu* de Queiroz 1998, 1999, 2007). This is shown by fairly deep divergence from its sister clade, the narrow-sense *africana* clade (V*a*), in addition to divergence from the narrow-sense *africana* clade in all other studied traits (Fig. 8) and widely allopatric distributions. However, the genetic divergence is considerably shallower than among the species proposed in this study, except *C. kidepoensis*, and it is also rather poorly differentiated in morphology, song and other behaviours from the taxa in the narrow-sense *africana* clade. In the study by Alström *et al.* (2023), *tropicalis* was almost as deeply diverged from *africana/transvaalensis* as the most recently diverged sympatric lark species pair. It clearly represents a borderline case with respect to species status, but we prefer to treat it as a subspecies of *C. africana s.s.*

The taxa *nigrescens* and *nyikae* are relatively deeply diverged in the multilocus analysis, but this is not supported by the SNP analyses. These taxa are data deficient, because we have examined only two specimens and photographs of three individuals of *nigrescens* and nine specimens and photographs of approximately seven individuals of *nyikae*, and only nine unique song strophes from *nigrescens* and six from *nyikae*. Their ranges are insufficiently known, and although their elevational preference (> 2,000 m a.s.l.) and the topography of the region suggest they are likely not to be in contact, we lack data. Future research is needed, but at present, we treat them as conspecific.

We also offer suggestions for vernacular names in English, for both the proposed new species and the previously recognised species with new circumscriptions (cf. Table 2), sequentially ordered based on the ASTRAL topology and following the principles of Fjeldså *et al.* (2020b):

(i) Highland lark for *C. kurrae*. The rather fragmented distribution across sub-Saharan West Africa to Sudan is restricted to areas of higher elevation than the surroundings, primarily between 500 and 1,600 m a.s.l.
(ii) Plains lark for *C. kabalii*. In those parts of the distribution for which we have more detailed information, the species is typically restricted to large, open watershed plains (Frank Willems in litt., who coined this name).
(iii) Plateau lark for *C. nigrescens*. This species occurs very localised on high plateaus, almost exclusively > 2,000 m a.s.l.: *C. n. nigrescens* on Kitulo Plateau and other areas in the north-western parts of the Kipengere Range, possibly including the extension Poroto Mountains; *C. n. nyikae* on the Nyika Plateau.
(iv) Rufous-naped lark for *C. africana* (*s.s.*). The name was used for *C. africana* (*s.l.*) and is suggested to remain for the rather widely distributed southern African clade that contains the nominate of the wider-sense species.
(v) Sentinel lark for *C. athi*. Based on their sentinel-like posture on various prominent perches, this name, coined by Brian Finch (in litt.), has received support among ornithologists within the distribution range of *C. athi*.
(vi) Somali lark for *C. somalica*. The species has been referred to alternately as the Somali lark or Somali long-billed lark, of which we favour the former, in order not to imply a close relationship to *Certhilauda* species.

(vii) Red-winged lark for *C. hypermetra* (*s.s.*). The name was used for *C. hypermetra* (*s.l.*) and is suggested to remain for the more widely distributed and well-known clade that contains the nominate of the wider-sense species.
(viii) Russet lark for *C. sharpii* (coined by Pamela C. Rasmussen). The upperparts of this species are coloured deep russet, with dark/black markings (cf. Figs 4, 8; Supporting Information, Fig. S2).
(ix) Kidepo lark for *C. kidepoensis*. The name refers to Kidepo Valley, where the taxon was discovered and which has also provided its scientific species epithet.

CONCLUSION

Our study highlights the lack of data in this species complex, especially on songs, other behaviours and distributions. We show that a range of strikingly diverse African taxa have previously been placed in a single umbrella species, *C. africana s.l.* Acknowledging this diversity and awarding appropriate taxonomic status, we suggest that conservation assessments are carried out urgently for some of the range-restricted and/or poorly known species. We hope that this paper will stimulate further interest in these birds, especially the ones that are still very poorly known (cf. Table 2), and deepen our understanding of various aspects of their biology and of biogeographical patterns and processes in Africa. The last word on the taxonomy of this complex has surely not been said.

SUPPORTING INFORMATION

Supplementary data is available at *Zoological Journal of the Linnean Society* online.

ACKNOWLEDGEMENTS

We are indebted to Mark Adams and Hein van Grouw and the Natural History Museum, Tring, UK, for access to specimens and toepad samples for DNA; to Matthew Medler and the Macaulay Library at Cornell University, USA, and Cheryl Tipp and the British Library's sound archive for access to sound recordings; to Nik Borrow, Françoise Dowsett-Lemaire, Gabriel B. Jamie, Michael Mills, David Moyer and Nigel Redman for additional sound recordings; to Nick Horrocks for managing DNA samples from Kenya; and to all sound recordists who have depositing their recordings with xeno-canto, Macaulay Library and the British Library; to Paul Sweet and the American Museum of Natural History, New York, for photographs of a *C. a. gomesi* specimen; to the Nigerian Montane Forest Project, Ngel Nyaki Forest Reserve for allowing E.B.E. to use Nigerian Montane Forest Project's (NMFP) sound recorder; to SciLifeLab for next-generation DNA sequencing; to the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, and the Swedish Research Council; the Swedish National Infrastructure for Computing (SNIC)/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure; and to the Research/Scientific Computing teams at The James Hutton Institute and the National Institute of Agricultural Botany for providing computational resources and technical support for the 'UK's Crop Diversity Bioinformatics HPC' (BBSRC grant BB/S019669/1). We thank Remco Bouckaert for advice on Snapper input specifications before those were publicly available. We are indebted to Françoise Dowsett-Lemaire and Robert J. Dowsett for information on type localities and references and comments on the first draft of the manuscript, and to the former also for checking the presence of mimicry in some sound recordings. We thank Dawie de Swardt, Adam S. Kennedy and Brian Finch for providing maps of rufous-naped and red-winged larks from their forthcoming guide to East African birds, and Brian Finch, Robert J. Dowsett, Gabriel Jamie, Pamela C. Rasmussen and Frank Willems for comments and suggestions for English names. For photographs and/or other assistance, we thank Nick Borrow, Frédéric Bacuez, Tomas Carlberg, Per Holmen, Conrad Marshall, David Moyer, Håkan Pohlstrand, Nigel Redman, Claire Spottiswoode, Stratton Hatfield and Washington Wachira, and we thank Beatriz Acuña Hidalgo for photograph editing. We thank an anonymous referee for feedback on previous versions of this manuscript.

FUNDING

This research was financially supported by the National Swedish Research Council (grant numbers 2015-04402 and 2019-04486), the Jornvall Foundation, the Carl Trygger Foundation (CTS 20:6) and Julian Francis (all to P.A.).

AUTHOR CONTRIBUTIONS

P.A., P.F.D. and M.S. conceived the study; P.A., D.E., H.K.N. and B.I.T. collected key DNA samples; M.I., U.O. and P.A. did the laboratory work for whole-genome sequencing; E.D.E. and M.S. conducted the bioinformatic analyses; M.S. performed all phylogenomic analyses; L.R. performed introgression analyses; Z.M. performed the multilocus analyses; P.A., D.E. and E.B.E. collected sound recordings; M.N. and P.A. took the song measurements; P.F.D. performed the song analyses; M.S., P.F.D. and P.A. recorded morphometrics; M.S. and P.F.D. performed morphometric analyses; P.A., M.S. and P.F.D. analysed plumage; P.A. and M.S. wrote the first draft of the manuscript. All authors commented on and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY

The sequences generated in this study have been submitted to GenBank under the accession numbers OR102883–OR102901. The SNP datasets (.vcf files) have been deposited to Zenodo, with the DOI 10.5281/zenodo.8060822.

REFERENCES

Alström P. Taxonomy of the *Mirafra assamica* complex. *Forktail* 1998;**13**:97–107.
Alström P. Tibetan lark (*Melanocorypha maxima*), version 1.0. In: del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E (eds), *Birds of the World*. Ithaca: Cornell Lab of Ornithology, 2020a. https://doi.org/10.2173/bow.tiblar1.01
Alström P. Mongolian lark (*Melanocorypha mongolica*), version 1.0. In: del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E (eds), *Birds of the World*. Ithaca: Cornell Lab of Ornithology, 2020b. https://doi.org/10.2173/bow.monlar1.01
Alström P. Oriental skylark (*Alauda gulgula*), version 1.0. In: del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E (eds), *Birds of the World*. Ithaca: Cornell Lab of Ornithology, 2020c. https://doi.org/10.2173/bow.orisky1.01

Alström P, Barnes KN, Olsson U *et al*. Multilocus phylogeny of the avian family Alaudidae (larks) reveals complex morphological evolution, non-monophyletic genera and hidden species diversity. *Molecular Phylogenetics and Evolution* 2013a;**69**:1043–56.

Alström P, Ericson PGP, Olsson U *et al*. Phylogeny and classification of the avian superfamily Sylvioidea. *Molecular Phylogenetics and Evolution* 2006;**38**:381–97.

Alström P, Gombobaatar SG. Mongolian short-toed lark (*Calandrella dukhunensis*), version 2.0. In: Billerman SM, Keeney BK (eds), *Birds of the World*. Ithaca: Cornell Lab of Ornithology, 2021. https://doi.org/10.2173/bow.sstlar4.02

Alström P, Mohammadi Z, Enbody ED *et al*. Systematics of the avian family Alaudidae using multilocus and genomic data. *Avian Research* 2023;**14**:100095. https://doi.org/10.1016/j.avrs.2023.100095

Alström P, Olsson U, Lei F. A review of the recent advances in the systematics of the avian superfamily Sylvioidea. *Chinese Birds* 2013b;**4**:99–131.

Alström P, Ranft R. The use of sounds in avian systematics, and the importance of bird sound archives. *Bulletin of the British Ornithologists' Club Supplement* 2003;**123A**:114–35.

Alström P, Sundev G. Mongolian short-toed lark *Calandrella dukhunensis*, an overlooked East Asian species. *Journal of Ornithology* 2021;**162**:165–77.

Alström P, van Linschooten J, Donald PF *et al*. Multiple species delimitation approaches applied to the avian lark genus *Alaudala*. *Molecular Phylogenetics and Evolution* 2021;**154**:106994. https://doi.org/10.1016/j.ympev.2020.106994

Andrews S. *FastQC: a Quality Control Tool for High Throughput Sequence Data*. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ash JS, Atkins J. *Birds of Ethiopia and Eritrea: an Atlas of Distribution*. London: Christopher Helm, 2009.

Ash JS, Miskell JE. *Birds of Somalia*. Robertsbridge: Pica Press, 1998.

Avise JC, Robinson TJ. Hemiplasy: a new term in the lexicon of phylogenetics. *Systematic Biology* 2008;**57**:503–7.

Baker L, Baker N. Rufous-naped lark. In: Baker L, Baker N (eds), *Tanzania Bird Atlas, Preview BETA*, 2014. http://tanzaniabirdatlas.net/maphtm/0612_nmap.htm

Beckman EJ, Benham PM, Cheviron ZA *et al*. Detecting introgression despite phylogenetic uncertainty: the case of the South American siskins. *Molecular Ecology* 2018;**27**:4350–67.

Berends MS, Luz CF, Friedrich AW *et al*. AMR: an R package for working with antimicrobial resistance data. *Journal of Statistical Software* 2022;**104**:1–31.

Bioacoustics Research Program. *Raven Pro. Interactive Sound Analysis Software, v. 1.5*. Ithaca: Cornell Lab of Ornithology, 2017. https://ravensoundsoftware.com

BirdLife International & Handbook of the Birds of the World. *Bird Species Distribution Maps of the World. Version 2020.1*, 2020. http://datazone.birdlife.org/species/requestdis

Bouckaert R, Vaughan TG, Barido-Sottani J *et al*. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 2019;**15**:e1006650.

Britton PL (ed). *Birds of East Africa: Their Habitat, Status and Distribution*. Nairobi: East Africa Natural History, 1980.

Bryant D, Bouckaert R, Felsenstein J *et al*. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* 2012;**29**:1917–32.

Butlin RK. Speciation by reinforcement. *Trends in Ecology & Evolution* 1987;**2**:8–13.

Campbell RW, Van Damme LM, Johnson SR *et al*. Eurasian skylark (*Alauda arvensis*), version 1.0. In: Billerman SM (ed), *Birds of the World*. Ithaca: Cornell Lab of Ornithology, 2020. https://doi.org/10.2173/bow.skylar.01

Charif RA, Waack AM, Strickman LM. *Raven Pro 1.4*. Ithaca: Cornell Lab of Ornithology, 2010.

Chen S, Zhou Y, Chen Y *et al*. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;**34**:i884–90.

Clancey PA. Miscellaneous taxonomic notes on African birds VII. *Durban Museum Novitates* 1956;**4**:273–91.

Clements JF, Schulenberg TS, Iliff MJ *et al*. *The eBird/Clements Checklist of Birds of the World: v2022*, 2022. https://www.birds.cornell.edu/clementschecklist/download

Colston PR. A new species of *Mirafra* (Alaudidae) and new races of the Somali Long-billed Lark *Mirafra somalica*, Thekla Lark *Galerida malabarica* and Malindi Pipit *Anthus melindae* from southern coastal Somalia. *Bulletin of the British Ornithologists' Club* 1982;**102**:106–14.

Cracraft J. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: Otte D, Endler JA (eds), *Speciation and its Consequences*. Sunderland: Sinauer Associates, 1989, 28–59.

Cracraft J. Species concepts in systematics and conservation biology – an ornithological viewpoint. In: Claridge MF, Dawah HA, Wilson MR (eds), *Species: the Units of Biodiversity*. London: Chapman & Hall, 1997, 325–39.

Cramp S. *The Birds of the Western Palearctic*, Vol. **V**. Oxford: Oxford University Press, 1988.

Danecek P, Auton A, Abecasis G *et al*.; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics* 2011;**27**:2156–8.

Danecek P, Bonfield JK, Liddle J *et al*. Twelve years of SAMtools and BCFtools. *GigaScience* 2021;**10**:giab008.

Da Rosa Pinto AA. Algumas formas novas para Angola e outras para a Ciência descobertas no distrito do Cuando-Cubango (Angola). *Bonner Zoologische Beiträge* 1968;**19**:280–8.

Darriba D, Taboada GL, Doallo R *et al*. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 2012;**9**:772.

Dean WRJ. Rufous-naped lark *Mirafra africana*. In: Hockey PAR, Dean WRJ, Ryan PG (eds), *Roberts Birds of Southern Africa*, 7th edn. Cape Town: The Trustees of the John Voelcker Bird Book Fund, 2005, 862–3.

Dean WRJ, Keith S. *Mirafra africana*. In: Keith S, Urban EK, Fry CH (eds), *The Birds of Africa, IV. Broadbills to Chats*. San Diego: Academic Press, 1992, 22–4.

De Juana E, Suárez F, Ryan P *et al*. Family Alaudidae (larks). In: del Hoyo J, Elliott A, Christie DA (eds), *Handbook of the Birds of the World*, Vol. **9**. Barcelona: Lynx Edicions, 2004, 496–601.

Del Hoyo J, Collar NJ. *HBW and BirdLife International Illustrated Checklist of the Birds of the World, Vol. 2: Passerines*. Barcelona: Lynx Edicions, 2016.

De Queiroz K. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In: Howard DJ, Berlocher SH (eds), *Endless Forms: Species and Speciation*. Oxford: Oxford University Press, 1998, 57–75.

De Queiroz K. The general lineage concept of species and the defining properties of the species category. In: Wilson RA (ed), *Species: New Interdisciplinary Essays*. Cambridge: MIT Press, 1999, 49–89.

De Queiroz K. Species concepts and species delimitation. *Systematic Biology* 2007;**56**:879–86.

Dickinson EC, Remsen JV Jr (eds). *The Howard and Moore Complete Checklist of the Birds of the World Fourth Edition. Vol. 2: Passerines*. Eastbourne: Aves Press. 2014.

Donald P, Alström P, Engelbrecht D. Possible mechanisms of substrate colour-matching in larks (Alaudidae) and their taxonomic implications. *Ibis* 2017;**159**:699–702.

Dowsett-Lemaire F, Dowsett RJ. Vocal mimicry in the lark *Mirafra hypermetra* as a possible species-isolating mechanism. *Bulletin of the British Ornithologists' Club* 1978;**98**:140–4.

Drovetski SV, Rakovic M, Semenov G *et al*. Limited phylogeographic signal in sex-linked and autosomal loci despite geographically, ecologically, and phenotypically concordant structure of mtDNA variation in the avian genus *Eremophila*. *PLoS One* 2014;**9**:e87570.

Drummond AJ, Ho SY, Phillips MJ *et al*. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 2006;**4**:e88.

Drummond AJ, Suchard MA, Xie D *et al*. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 2012;**29**:1969–73.

Edwards SV, Potter S, Schmitt CJ *et al.* Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proceedings of the National Academy of Sciences of the United States of America* 2016;**113**:8025–32.
Fjeldså J, Alström P, Olsson U *et al.* Superfamily Sylvioidea, the Old World warblers and their allies. In: Fjeldså J, Ericson PGP, Christidis L (eds), *The Largest Avian Radiation: the Evolution of Perching Birds, or the Order Passeriformes*. Barcelona: Lynx Edicions, 2020a, 191–236.
Fjeldså J, Christidis L, Ericson PGP *et al.* An updated classification of passerine birds. In: Fjeldså J, Ericson PGP, Christidis L (eds), *The Largest Avian Radiation: the Evolution of Perching Birds, or the Order Passeriformes*. Barcelona: Lynx Edicions, 2020b, 45–64.
Freed D, Aldana R, Weber JA *et al.* The Sentieon Genomics Tools – a fast and accurate solution to variant calling from next-generation sequence data. *BioRxiv* 2017:115717. https://doi.org/10.1101/115717. Preprint: not peer reviewed.
Fregin S, Haase M, Olsson U *et al.* New insights into family relationships within the avian superfamily Sylvioidea (Passeriformes) based on seven molecular markers. *BMC Evolutionary Biology* 2012;**12**:157.
Ghorbani F, Aliabadian M, Olsson U *et al.* Mitochondrial phylogeography of the genus *Eremophila* confirms underestimated species diversity in the Palearctic. *Journal of Ornithology* 2020a;**161**:297–312.
Ghorbani F, Aliabadian M, Zhang R *et al.* Densely sampled phylogenetic analyses of the lesser short-toed lark *Alaudala rufescens* — sand lark *A. raytal* species complex (Aves, Passeriformes) reveal cryptic diversity. *Zoologica Scripta* 2020b;**49**:427–39.
Gill F, Donsker D, Rasmussen P (eds). *IOC World Bird List (v. 13.1)*, 2023. https://doi.org/10.14344/IOC.ML.13.1
Guillaumet A, Crochet PA, Godelle B. Phenotypic variation in *Galerida* larks in Morocco: the role of history and natural selection. *Molecular Ecology* 2005;**14**:3809–21.
Guillaumet A, Crochet PA, Pons JM. Climate-driven diversification in two widespread *Galerida* larks. *BMC Evolutionary Biology* 2008;**8**:32.
Guillaumet A, Pons JM, Godelle B *et al.* History of the crested lark in the Mediterranean region as revealed by mtDNA sequences and morphology. *Molecular Phylogenetics and Evolution* 2006;**39**:645–56.
Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research* 2013;**41**:e129.
Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999;**41**:95–8.
Hoang DT, Chernomor O, von Haeseler A *et al.* UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 2018;**35**:518–22.
Hudson EJ, Price TD. Pervasive reinforcement and the role of sexual selection in biological speciation. *Journal of Heredity* 2014;**105**:821–33.
Irestedt M, Thörn F, Müller I *et al.* A guide to avian museomics: insights gained from resequencing hundreds of avian study skins. *Molecular Ecology Resources* 2022;**22**:2672–84.
Junier T, Zdobnov EM. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* 2010;**26**:1669–70.
Kalyaanamoorthy S, Minh BQ, Wong TKF *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017;**14**:587–9.
Kennedy AS, Finch BW. *Birds of East Africa: an Identification Guide*. Princeton: Princeton University Press, In prep.
Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013;*1303.3997 [q-bio.GN]*. https://doi.org/10.48550/arXiv.1303.3997. Preprint: not peer reviewed.
Malinsky M, Matschiner M, Svardal H. Dsuite - fast *D*-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources* 2021;**21**:584–95.
Malinsky M, Svardal H, Tyers AM *et al.* Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology and Evolution* 2018;**2**:1940–55.
Mason NA, Pulgarin P, Cadena CD *et al. De novo* assembly of a high-quality reference genome for the horned lark (*Eremophila alpestris*). *G3: Genes, Genomes, Genetics* 2020;**10**:475–8.
Mayr E. *Systematics and the Origin of Species*. Cambridge: Harvard University Press, 1942.
Mayr E. *Animal Species and Evolution*. Cambridge: Harvard University Press, 1963.
Mayr E. *Principles of Systematic Zoology*. New York: McGraw-Hill, 1969.
McKenna A, Hanna M, Banks E *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 2010;**20**:1297–303.
Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* 2010;**2010**:pdb.prot5448.
Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)* 2010;**2010**:1–8.
Nguyen LT, Schmidt HA, von Haeseler A *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 2015;**32**:268–74.
Olsson U, Alström P, Ericson PGP *et al.* Non-monophyletic taxa and cryptic species—evidence from a molecular phylogeny of leaf-warblers (*Phylloscopus*, Aves). *Molecular Phylogenetics and Evolution* 2005;**36**:261–76.
Ortiz EM. Vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. *Zenodo* 2019:2540861. https://doi.org/10.5281/zenodo.2540861. Online deposition Jan15, 2019.
Päckert M. Song: the learned language of three major bird clades. In: Tietze DT (ed), *Bird Species: How They Arise, Modify and Vanish*. Cham: Springer Open, 2018, 75–94.
Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;**35**:526–8.
Patterson N, Moorjani P, Luo Y *et al.* Ancient admixture in human history. *Genetics* 2012;**192**:1065–93.
Pease JB. Why phylogenomic uncertainty enhances introgression analyses. *Molecular Ecology* 2018;**27**:4347–9.
Peters JL. Family Alaudidae. In: Mayr E, Greenway JC Jr (eds), *Check-list of Birds of the World*, Vol. **IX**. Cambridge: Museum of Comparative Zoology, 1960, 3–80.
Price T. *Speciation in Birds*. Greenwood Village: Roberts and Company, 2008.
R Core Team. *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2021. https://www.R-project.org/
Rambaut A. FigTree, v. 1.4.3, 2016. http://tree.bio.ed.ac.uk/software/figtree/.
Rambaut A. *FigTree, Version 1.7.1*, 2018. http://tree.bio.ed.ac.uk/software/figtree/
Rambaut A, Drummond AJ. TreeAnnotator, v. 2.2.1, 2015. http://www.beast2.bio.ed.ac.uk.
Rancilhac L, Irisarri I, Angelini C *et al.* Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders. *Molecular Phylogenetics and Evolution* 2021;**155**:106967. https://doi.org/10.1016/j.ympev.2020.106967
Rasmussen PC, Anderton JC. *Birds of South Asia: The Ripley Guide*, 2nd revised edn, Vol. **2**. Barcelona: Lynx Edicions, Smithsonian Institution and Michigan State University, 2012.
Roche J. Recherches écologiques et systématiques sur les oiseaux de la République de Somalie (partie ex-italienne). *Monitore Zoologico Italiano* 1966;**1**:44–79.
Ryan P. Red-winged lark (*Mirafra hypermetra*), version 1.0. In: del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E (eds), *Birds of the World*. Ithaca: Cornell Lab of Ornithology, 2020a. https://doi.org/10.2173/bow.rewlar1.01
Ryan P. Somali long-billed lark (*Mirafra somalica*), version 1.0. In: del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E (eds), *Birds of the World*. Ithaca: Cornell Lab of Ornithology, 2020b. https://doi.org/10.2173/bow.slblar1.01
Ryan P. Ash's lark (*Mirafra ashi*), version 1.0. In: del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E (eds), *Birds of the World*. Ithaca: Cornell Lab of Ornithology, 2020c. https://doi.org/10.2173/bow.ashlar1.01
Ryan PG, Bloomer P. The long-billed lark complex: a species mosaic in south-western Africa. *The Auk* 1999;**116**:194–208.

Ryan P, del Hoyo J, Collar N *et al*. Rufous-naped lark (*Mirafra africana*), version 1.0. In: Billerman SM, Keeney BK, Rodewald PG, Schulenberg TS (eds), *Birds of the World*. Ithaca: Cornell Lab of Ornithology, 2020. https://doi.org/10.2173/bow.runlar1.01

Ryan PG, Hood I, Bloomer P *et al*. Barlow's lark: a new species in the Karoo lark *Certhilauda albescens* complex of southwest Africa. *Ibis* 1998;**140**:605–19.

Sangster G. Integrative taxonomy of birds: the nature and delimitation of species. In: Tietze DT (ed), *Bird Species: How They Arise, Modify and Vanish*. Cham: Springer Open, 2018, pp. 9–37.

Shirihai H, Svensson L. *Handbook of Western Palearctic Birds*. London: Helm, 2018.

Shirihai H, Belmaker A, Feldstein-Farkash T *et al*. The Mongolian short-toed lark *Calandrella dukhunensis* in Israel: an overlooked species in the Western Palearctic? *British Birds* 2023;**116**:252–9.

Spottiswoode CN, Olsson U, Mills MSL *et al*. Rediscovery of a long-lost lark reveals the conspecificity of endangered *Heteromirafra* populations in the Horn of Africa. *Journal of Ornithology* 2013;**154**:813–25.

Stange M, Sánchez-Villagra MR, Salzburger W *et al*. Divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian isthmus. *Systematic Biology* 2018;**67**:681–99.

Stervander M, Alström P, Olsson U *et al*. Multiple instances of paraphyletic species and cryptic taxa revealed by mitochondrial and nuclear RAD data for *Calandrella* larks (Aves: Alaudidae). *Molecular Phylogenetics and Evolution* 2016;**102**:233–45.

Stervander M, Hansson B, Olsson U *et al*. Molecular species delimitation of larks (Aves: Alaudidae), and integrative taxonomy of the genus *Calandrella*, with the description of a range-restricted African relic taxon. *Diversity* 2020;**12**:428. https://doi.org/10.3390/d12110428

Stoltz M, Baeumer B, Bouckaert R *et al*. Bayesian inference of species trees using diffusion models. *Systematic Biology* 2021;**70**:145–61.

Taylor SA, Larson EL. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution* 2019;**3**:170–7.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 1994;**22**:4673–80.

Weir JT, Schluter D. Calibrating the avian molecular clock. *Molecular Ecology* 2008;**17**:2321–2328.

White CMN. New races of lark, pipit, and cordon bleu from northern Rhodesia, Angola, and the Belgian Congo. *Bulletin of the British Ornithologists' Club* 1944;**65**:5–7.

Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer, 2009.

Zhang C, Rabiee M, Sayyari E *et al*. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 2018;**19**:153.

Zhang D, Rheindt FE, She H *et al*. Most genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow. *Systematic Biology* 2021;**70**:961–75.